Adaptive Risk Management

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Dr. rer. pol.)
im Fach Wirtschaftswissenschaft

eingereicht an der
Wirtschaftswissenschaftliche Fakultät
Humboldt-Universität zu Berlin

von
Frau M.Sc. Ying Chen
geboren am 08.12.1975 in Kaifeng, VR China

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Wirtschaftswissenschaftliche Fakultät:
Prof. Oliver Günther, Ph.D.

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Prof. Dr. Vladimir Spokoiny

eingereicht am:	15. Oktober 2006
Tag der mündlichen Prüfung:	9. Januar 2007

Abstract

Over recent years, study on risk management has been prompted by the Basel committee for the requirement of regular banking supervisory. There are however limitations of many risk management methods: 1) covariance estimation relies on a time-invariant form, 2) models are based on unrealistic distributional assumption and 3) numerical problems appear when applied to high-dimensional portfolios.

The primary aim of this dissertation is to propose adaptive methods that overcome these limitations and can accurately and fast measure risk exposures of multivariate portfolios. The basic idea is to first retrieve out of high-dimensional time series stochastically independent components (ICs) and then identify the distributional behavior of every resulting IC in univariate space. To be more specific, two local parametric approaches, local moving window average (MWA) method and local exponential smoothing (ES) method, are used to estimate the volatility process of every IC under the heavy-tailed distributional assumption, namely ICs are generalized hyperbolic (GH) distributed. By doing so, it speeds up the computation of risk measures and achieves much better accuracy than many popular risk management methods.

Keywords:
Risk management, Heavy-tailed distribution, Local parametric methods, High-dimensional data analysis

Zusammenfassung

In den vergangenen Jahren ist die Untersuchung des Risikomanagements vom Baselkomitee angeregt, um die Kredit- und Bankwesen regelmäßig zu aufsichten. Für viele multivariate Risikomanagementmethoden gibt es jedoch Beschränkungen von: 1) verlässt sich die Kovarianzschätzung auf eine zeitunabhängige Form, 2) die Modelle beruhen auf eine unrealistischen Verteilungsannahme und 3) numerische Problem, die bei hochdimensionalen Daten auftreten.
Es ist das primäre Ziel dieser Doktorarbeit, präzise und schnelle Methoden vorzuschlagen, die diesen Beschränkungen überwinden. Die Grundidee besteht darin, zuerst aus einer hochdimensionalen Zeitreihe die stochastisch unabhängigen Komponenten (IC) zu extrahieren und dann die Verteilungsparameter der resultierenden IC beruhend auf eindimensionale Heavy-Tailed Verteilungsannahme zu identifizieren. Genauer gesagt werden zwei lokale parametrische Methoden verwendet, um den Varianzprozess jeder IC zu schätzen, das lokale Moving Window Average (MVA) Methode und das lokale Exponential Smoothing (ES) Methode. Diese Schätzungen beruhen auf der realistischen Annahme, dass die IC Generalized Hyperbolic (GH) verteilt sind. Die Berechnung ist schneller und erreicht eine höhere Genauigkeit als viele bekannte Risikomanagementmethoden.

Schlagwörter:
Risikomanagement, Heavy-Tailed Verteilung, Lokale parametrische Methoden, Hochdimensionale Datenanalyse

Contents

List of Figures

List of Tables

Part I

Basic Concepts

Chapter 1

Introduction

After the breakdown of the Bretton Woods fixed exchange rate system in 1971, financial markets have become much more volatile than before. The following boom of financial derivatives accelerated the turbulence of the markets. Under such a situation, Basel committee on banking regulations and supervisory was founded by central-bank governors of the Group of ten countries in 1974. The main goals of the Basel committee are to secure capital adequacy and control market risks of financial institutions. Despite the careful and strong regulatory, losses in trading financial instruments astonished the world due to the suddenness and the extremely large amount. For example, loss in trading financial derivatives totaled 28 billion US Dollar from 1987 to 1988, see [Jor01]. To alleviate the extreme market risks, Basel accord has asked financial institutions, mainly banks, to deposit capital of risk assets as risk charge. The standard rule has changed from the beginning "8% rule", i.e. capital reserved should be larger or equal to 8% of the risk-weighted assets, to the recent allowance of using "internal model", i.e. risk charge is decided based on a verified quantitative model used by the financial institution. Unfortunately, extreme losses are continuously observed in the market and tend to involve even larger amount of capital than before. Merely the loss of Barings has an amount of 1330 million US Dollar in trading stock index futures in 1995. These observations arise the questions: Whether the popular risk management models are appropriate for measuring risk exposures in the more and more sophisticated financial markets and How the risk management models can be improved.

It is important to investigate features of financial series before measuring its risk. Financial series shares a number of stylized facts such as non-stationary property of stock prices, volatility clustering fact and heavy-tailed distribution, see [Fam65] and [Pag96]. We here demonstrate these features on the base of the German stock Allianz from 1988/01/04 to 1996/12/30. Given

the observed stock price $s(t)$, the Dickey-Fuller test is to check whether that the stock price follows a random walk $s(t) = c+s(t-1)+\varepsilon_s(t)$ with some constant c and stochastic innovations $\varepsilon_s(t)$. We refer to [DF79] for details of this unit root test. With a value of -3.22 we can not reject the non-stationarity hypothesis at the 5% level since the corresponding critical value is -3.41. The test w.r.t. log-returns $x(t) = \log\{s(t)/s(t-1)\}$, on the contrary, rejects the non-stationarity hypothesis. For notational simplification, we use return to express log-return in this thesis. Notice that the rejection of the unit root test only states that the first order difference of the stock price does not follow random walk. The rejection is not sufficient to say that the return process is stationary, i.e. with at least the same mean and variance. In fact, it is observed that variance is changing through time. To be more specific, large returns tend to be followed by large returns, of either sign, and small changes tend to be followed by small changes. This phenomenon was first observed by [Man63] and is famous as "volatility clustering" feature. It is displayed in Figure 1.1, where the series changes between volatile scene and relative quiet scene through time. It is rational to surmise that a large variance exists in the volatile period and a small variance is for the quiet period. Although the variance of the return process is time varying, it is supposed that the standardized returns $x(t)/\ Var[x(t)]$ have a time-homogeneous variance with expected value of 1. Furthermore, two empirical high order moments, skewness and kurtosis, of the standardized returns are with values of -0.177 and 12.077. These values indicate that the return process has an asymmetric distribution and heavy tails relative to the Gaussian random variable with skewness $= 0$ and kurtosis $= 3$.

In accordance with these empirical features, financial risks are typically mapped into a stochastic framework:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t), \tag{1.1}$$

where $x(t) \in \mathbb{R}^d$ are risk factors, e.g. returns of a d-dimensional portfolio, the covariance $\Sigma_x(t)$ is time-dependent and will be filtered from the past observations $\{x(1), \cdots, x(t-1)\}$. The stochastic innovations $\varepsilon_x(t) \in \mathbb{R}^d$ are assumed to be independent and identical distributed (i.i.d.) with $\mathsf{E}[\varepsilon_{x_j}(t)|\mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\varepsilon^2_{x_j}(t)|\mathcal{F}_{t-1}] = 1$ for $j = 1, \cdots, d$. The popular risk measures such as value at risk (VaR) and expected shortfall (ES) are calculated based on the estimated joint density of the risk factors $x(t)$, therefore, the practical risk management decisions inherently rely on out-of-sample covariance forecasts and innovations' distributional identification, indicated by (1.1).

Many methodologies have been contributed to deal with these two tasks. In spite of theoretical and empirical achievements, they are often not uniformly applicable in risk management due to three limitations, see [Bol01].

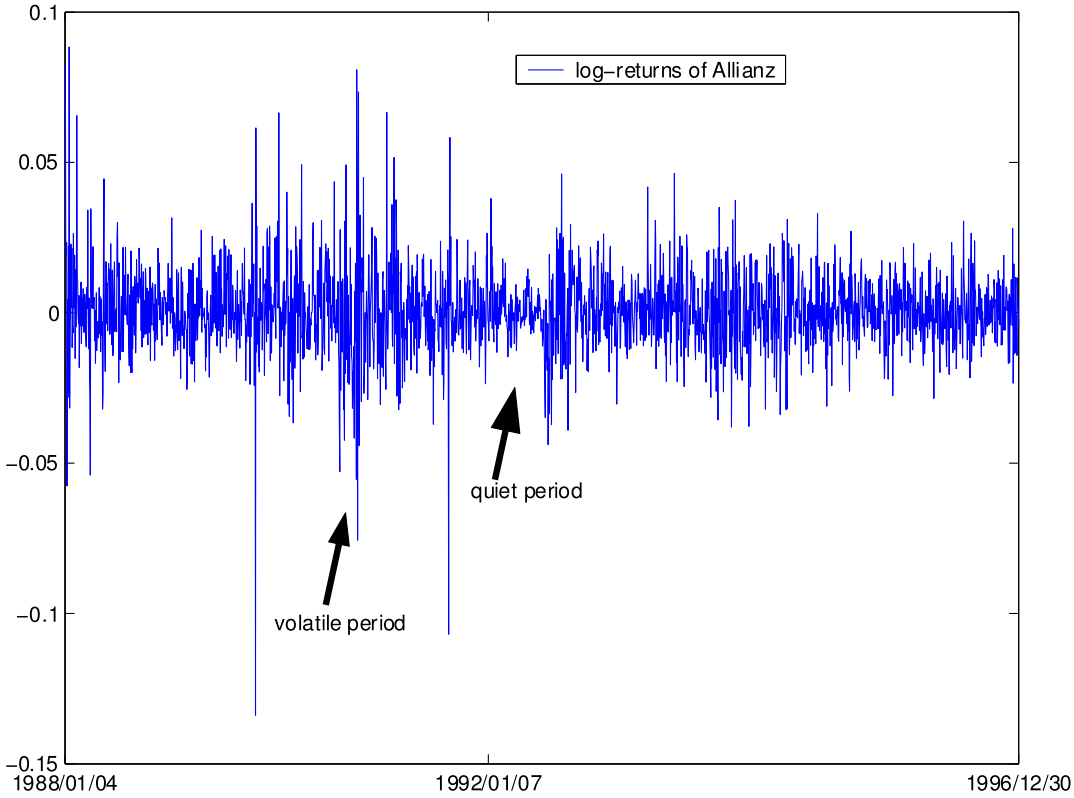

Figure 1.1: Log-returns of the German stock Allianz from 1988/01/04 to 1996/12/30. Data source: FEDC (http://sfb649.wiwi.hu-berlin.de)

1. **Fixed form in covariance estimation**: The covariance estimation methods normally rely on a fixed parametric regression form, namely the form is time-invariant. Such a static assumption is numerically tractable but may induce large estimation errors. Since it often observes structure shifts in the markets, which are driven by e.g. policy adjustments or economic changes. The static form has less flexibility to react to these shifts, and therefore, is weak to match covariance movement in a meaningful way.

2. **Gaussian distributional assumption**: The innovations are often assumed to be Gaussian distributed. This assumption gives fast and explicit results in the calculation, but the Gaussian distribution has relatively lighter tails than the empirical one. In risk management, risk measures are however calculated based on the tail part of the identified distribution. Therefore this Gaussian assumption costs too much in losing accuracy.

3. **Numerical problem due to high-dimensionality**: Last but not least, it is difficult to implement the popular models in high-dimensional analysis. Although large dimensional portfolios are actively traded by many financial institutions. The problem is that an appropriate dimensional reduction method is lacking in risk management.

Now let us give more details on these three limitations and briefly explain our ideas. First, two covariance estimation methods, the moving window average (MWA) method and the exponential smoothing (ES) method, are in particular desirable in practice since they are successful in reflecting the volatility clustering fact of financial series with simple forms:

$$\begin{aligned}
\text{MWA:}\quad \Sigma_x(t) &= \frac{1}{M}\left\{\sum_{m=0}^{M} x(t-m-1)x^\top(t-m-1)\right\}, \quad M \le t-2 \\
\text{ES:}\quad \Sigma_x(t) &= \left\{\sum_{m=0}^{\infty} \eta^m x(t-m-1)x^\top(t-m-1)\right\} \Big/ \left\{\sum_{m=0}^{\infty} \eta^m\right\}, \\
&\quad \eta \in [0,1]
\end{aligned}$$

Notice that the ARCH and GARCH models in [Eng82] and [Bol86] can be considered as the variations of the ES method:

$$\begin{aligned}
\Sigma_x(t) &= \omega + \alpha x(t-1)x^\top(t-1) + \beta\Sigma_x(t-1) \\
&= \frac{\omega}{1-\beta} + \alpha\sum_{m=0}^{\infty}\beta^m x(t-m-1)x^\top(t-m-1)
\end{aligned}$$

time period	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
1988/01/04-1989/10/13	8.63e-06 (6.36e-06)	0.07 (0.03)	0.87 (0.05)
1989/10/13-1991/08/07	6.54e-06 (2.95e-06)	0.17 (0.07)	0.61 (0.12)
1988/01/04-1991/08/07	1.61e-05 (6.93e-06)	0.12 (0.04)	0.83 (0.04)

Table 1.1: ML estimates of the GARCH(1,1) model on the base of the German stock Allianz. The standard deviation of the estimates are reported in parentheses.

To implement these estimation methods, one needs to choose the value of the smoothing parameter M or η. The potential problem is that large estimation errors may appear due to the fixed parameter over a long time period. This limitation is illustrated in the univariate case by estimating the volatility of the Allianz stock. Over the nine years (1988 to the end of 1996), many policy adjustments and important events occurred in economic life. For example, large negative returns were observed in the US and European stock markets on 13 October 1989. The Allianz for example dropped over 13% on the date. Luckily, the downward movement neither destroyed the confidence of investors nor became a stock crash. Even through, the structure of this stock changed. Table 1.1 shows this case, where the returns of the Allianz before and after 13 October 1989 are respectively modelled in the GARCH(1,1) setup, $\sigma^2(t) = \omega + \alpha x^2(t-1) + \beta\sigma^2(t-1)$. We refer to [BW92] for maximum likelihood estimation of the parameters. In the estimation we consider the same sample size for the two subsets, by cutting the second sample on 1991/08/07. The maximum likelihood estimate (MLE) of the involved parameters, as expected, are distinct for the two subsets. The parameter β, for instance, changes from 0.87 before the drop to 0.61 after the drop. The standard deviation of the estimates are put in parentheses. Figure 1.2 details the volatility estimates w.r.t. the two subsets and these estimates based on the whole concerned time period as well. It shows that the estimated process is smoother before the drop than after given the larger smoothing parameter $\hat{\beta} = 0.87$. Remember a large smoothing parameter here corresponds to a low variation of estimates since more historical observations are used in the estimation. Furthermore, the volatility estimates based on the whole time period from 1988/01/04 to 1991/08/07 present different values from those based on the two small subsets due to different values of the smoothing parameter.

Given the example, it is interesting to ask which MLEs of β we should use for the second sample, $\hat{\beta} = 0.83$ or $\hat{\beta} = 0.61$? [MS00] find that long range dependence effect is due to structural changes in the data and the volatility

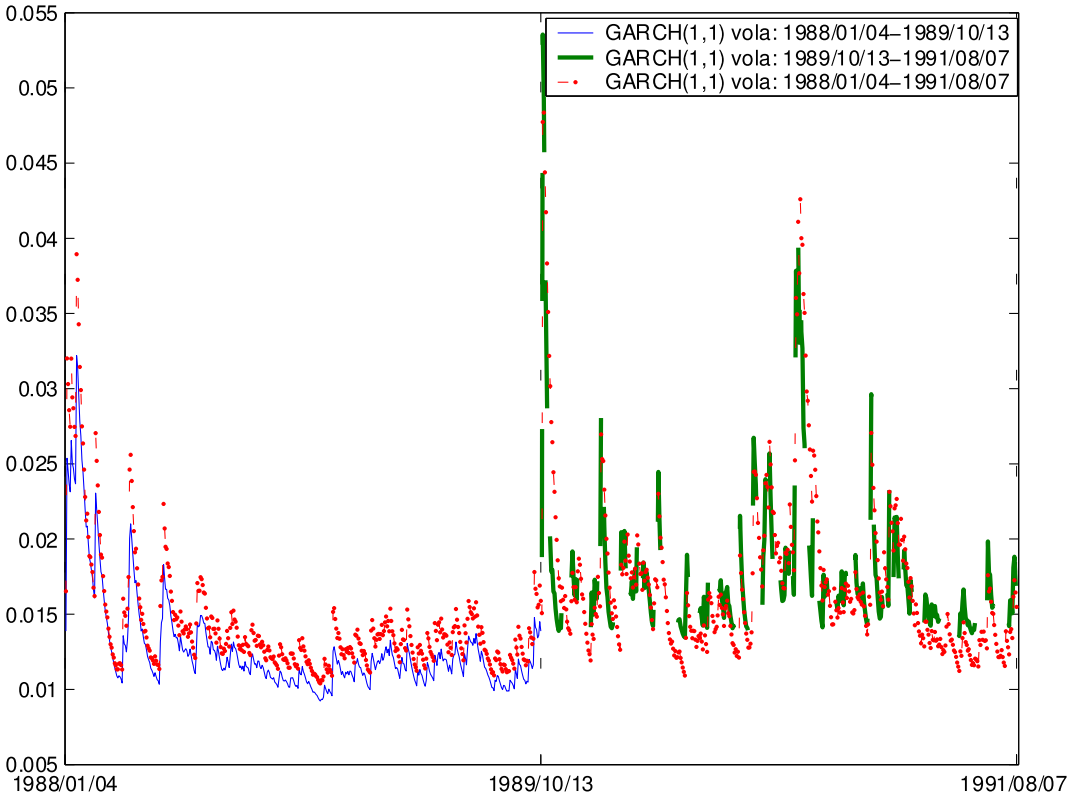

Figure 1.2: Volatility estimates of the German stock Allianz over three time periods: $\sigma^2(t) = \omega + \alpha x^2(t-1) + \beta\sigma^2(t-1)$. Data source: FEDC (http://sfb649.wiwi.hu-berlin.de)

clustering feature can be described by a locally stationary process. To be detailed, only the recent observations are important and useful for volatility estimation, indicating to choose a small value of smoothing parameter. A more plausible way is to adaptive the smoothing parameter through time. By doing so, the covariance estimation methods alleviate the potential misspecified problem and enhance the accuracy of estimation. [MS04a] present a local constant model, assuming that volatility changes little over a short interval and can be estimated using the local moving window average. To be more specific, the smoothing parameter M is time-dependent and individually selected for every time point. The consequent study of [CHJ05] extend the theory from the Gaussian distributional assumption to a realistic distributional framework by assuming that the series are generalized hyperbolic (GH) distributed. In the recent, [CS06] present the local exponential smoothing method by adaptively choosing the smoothing parameter η through time, by which the stochastic financial series are either Gaussian or GH distributed.

Besides the limitation of the covariance estimation, the mainstay of many risk management models is the Gaussian distributional assumption, e.g. the RiskMetrics product introduced by JP Morgan in 1994. This distributional assumption is based on the belief that the sum of a large amount of returns asymptotically converges to the Gaussian distribution, i.e. the central limit theory. In the Gaussian framework with an estimate $\hat{\Sigma}_x(t)$ of $\Sigma_x(t)$, the standardized returns $\hat{\varepsilon}_x(t) = \hat{\Sigma}_x^{(-1/2)}(t)x(t)$ are asymptotically independent and the joint distributional behavior can be easily measured by the marginal distributions. Another reason for assuming Gaussian distribution is that the VaR at 95% confidence level based on the Gaussian distributional assumption is almost identical to that with a more realistic heavy-tailed distribution such as GH distribution, see [JJ02]. Nevertheless, the Gaussian assumption is not appropriate for the modern risk management. The conditional Gaussian marginal distributions and the resulting joint Gaussian distribution are first at odds with empirical facts, i.e. financial series are heavy-tailed distributed. Even the highly diversified portfolios by trading high-dimensional financial instruments, are only closer to the Gaussian distribution than any individual returns, but they still deviate from the target assumption. Moreover, as the financial markets become more and more complex than before, VaRs with higher confidence levels, such as 99% level, have drawn the attention of risk analysts. These values are quite different in the Gaussian-based models and e.g. GH-based models.

Figure 1.3 demonstrates the effect of the distributional assumptions for two real data sets, the Allianz stock and a DAX portfolio from 1988/01/04 to 1996/12/30. The DAX is the leading index of Frankfurt stock exchange and a 20-dimensional hypothetic portfolio with a static trading strategy

$b(t) = (1/20, \cdots, 1/20)^\top$ is considered. The portfolio returns $r(t) = b(t)^\top x(t)$ are analyzed in the univariate version of (1.1). Notice that this simplification is possible in practice, but it often suffers from low accuracy of calculation. Suppose now that the two return processes have been properly standardized, by using the local exponential smoothing method. The standardized returns are empirically heavy-tailed distributed, indicated by the sample kurtoses 12.07 for the Allianz and 22.38 for the portfolio respectively. Three density estimations under the GH, Gaussian and Student-t with degrees of freedom 6 distributional assumptions are depicted in the figure. In the density comparison, the logarithmic density estimates using the nonparametric kernel estimation are considered as benchmark. The comparison w.r.t. the Allianz stock shows that the GH estimates are most close to the benchmark among the others. The Student-$t(6)$ has been recommended in practice due to its heavy tailedness. It however displays heavier tails relative to the benchmark, and the Gaussian estimates, on the contrary, present lighter tails. The similar result is observed w.r.t. the DAX portfolio. It is rational to surmise that the risk management methods under the Gaussian and $t(6)$ distributional assumptions generate low accurate results. This comparison motivates us to rely on the GH distributional assumption in the analysis.

Compared to the first two limitations in practice, the largest challenge of risk management is due to high-dimensionality of real portfolios. For example, many covariance estimation methods are really computationally demanding, even with static form. For example, the multivariate GARCH is recommended to estimate covariance matrix due to the good performance of its univariate version. The constant conditional correlation (CCC) model proposed by [Bol90] and the subsequent dynamic conditional correlation (DCC) model proposed by [Eng02], [ES01] have been considered as fast estimation methods, by which the covariance matrix is approximated by product of a diagonal matrix and a correlation matrix: $\Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top$. It reduces the number of unknown parameters relative to the former multivariate GARCH estimation model, the BEKK specification proposed by [EK95b]. Despite the appealing dimensional reduction, the mentioned estimation methods are still time consuming and numerically difficult to handle as really high-dimensional series, if e.g. a dimension $d > 10$, is considered, see [HHS03]. In addition, these estimation methods rely on the questionable Gaussian distributional assumption to ensure the independence of the resulting standardized returns. Otherwise, the distributional identification under a realistic assumption, such as the multivariate GH distribution with at least $4d$ parameters, involves once again numerical problem.

The primary aim of this thesis is to introduce fast and accurate risk management models for measuring portfolio risk exposure. As discussed before,

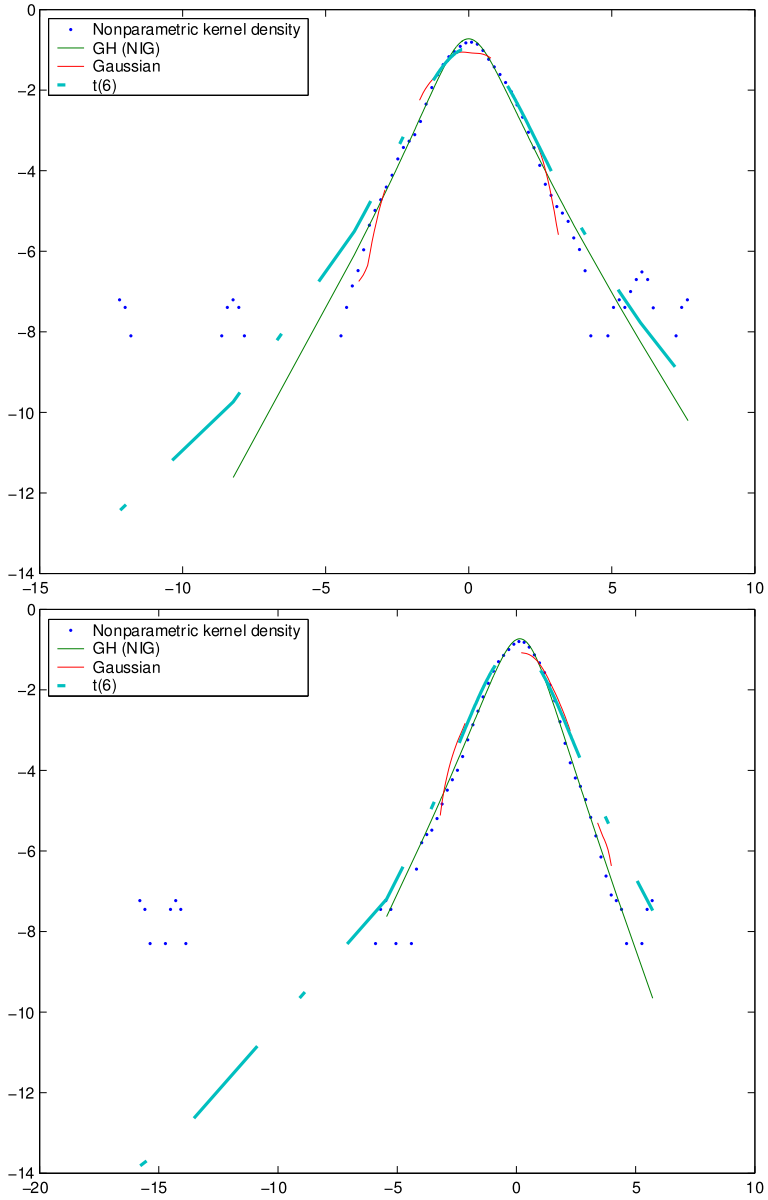

Figure 1.3: Density comparisons of the standardized returns in log scale based on the Allianz stock (top) and the DAX portfolio (bottom) with a static weight $b(t) = \text{unit}(1/20)$. Time interval: 1988/01/04 - 1996/12/30. The nonparametric kernel density is considered as benchmark. The GH distributional parameters are respectively $\text{GH}(-0.5, 1.01, 0.05, 1.11, -0.03)$ for the Allianz and $\text{GH}(-0.5, 1.21, -0.21, 1.21, 0.24)$ for the DAX portfolio. Data source: FEDC (sfb649.wiwi.hu-berlin.de).

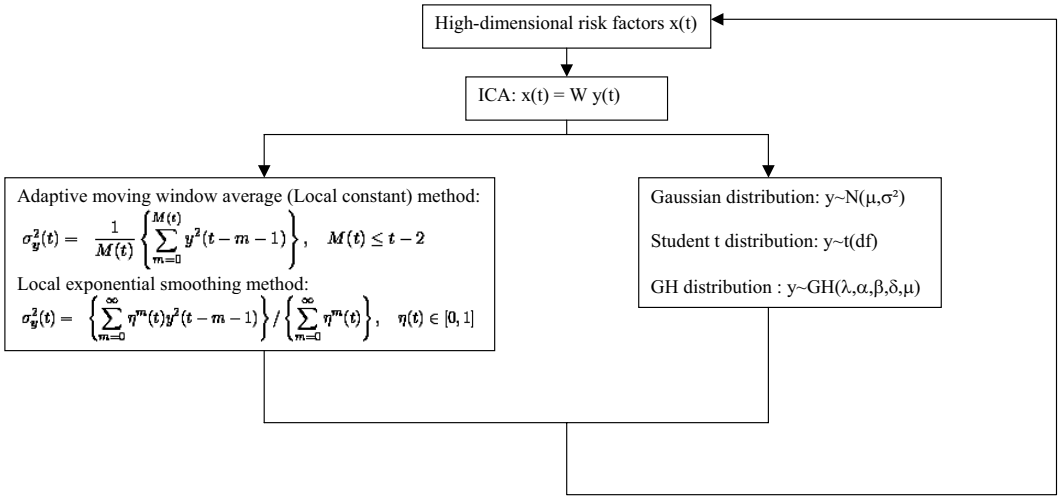

Figure 1.4: Procedure of adaptive risk management models.

the models should deal with high-dimensionality in an easy way, estimate covariance in a flexible way and identify stochastic behavior of financial series based on a realistic assumption. The general idea is illustrated in Figure 1.4. We first implement independent component analysis (ICA) method to achieve ICs. We then adaptively estimate the variance process of each resulting IC in a univariate space. The distributional behavior of each IC is identified in the GH framework. Thanks to the independence property of the ICs and the assumed linear relation in the ICA, it is easy to approximate the covariance and further the joint distribution of the original series. The second aim is to evaluate the proposed methods with the popular risk management models in the market. This thesis is organized as follows.

Chapter 2 introduces concepts of risk analysis and discuss the pros and cons of several popular risk management models based on the empirical features of return, volatility and more.

The following two chapters focus on the adaptive risk management methods given univariate time series. Chapter 3 introduces the GHADA risk management model based on the paper of [CHJ05], by which the local constant model is applied to estimate volatility under the GH distributional assumption. Compared to the Gaussian-based risk management model, the GHADA delivers very accurate results. Chapter 4 introduces the model based on the local exponential smoothing method. Adaptive methods are used to choose local smoothing parameter in the volatility estimation, see [CS06]. The qual-

ity of estimation is to a great extent enhanced by using the local methods. More important, both methods are applicable under the GH distributional assumption.

These proposed methods can be easily applied in high-dimensional risk analysis. Chapter 5 and Chapter 6 present multivariate risk management methods based on the papers of [CHS6a] and [CHS6b], by which the high-dimensional risk factors are first converted to ICs through a linear transformation. After that, the marginal distributional behavior of the ICs are measured by the proposed univariate methods. The quantile of the portfolio returns is estimated using Monte Carlo simulation and the fast Fourier transformation (FFT) technique respectively. Both methods produce nice results in the comparison with several popular risk management methods.

Chapter 2

Risk analysis

Sound risk management system has of great importance, since a large devaluation in the financial market is often followed by economic depression and bankruptcy of credit system. On the Monday, October 19 1987, for example, the Dow Jones industrial dropped by over 500 points. The worldwide stock trading markets suffered a similar devaluation. The stock crash destroyed the confidence of investors and consequently caused economic depression. For this reason, it is necessary to measure and control risk exposures using accurate methods. In this chapter, we first classify risks to different categories and introduce two popular risk measures. We then discuss the meaningfulness and desirability of measuring risks from the viewpoints of regulatory, internal supervisory and investors. Finally, we present several risk management methods widely used in the market and briefly describe the adaptive risk management methods.

2.1 Risk: classification and definition

Risks have many sources. Basel committee on banking supervisory classifies financial risk into market risk, credit risk, liquidity risk, operational risk and legal risk. In the following, we give brief definitions on these kind of risks.

- **Market risk**: It arises from the uncertainty due to changes in market prices and rates such as share prices, foreign exchange rates and interest rates, the correlations among them and their levels of volatility, see [Jor01]. The market risk is the main risk source and has a great negative influence on the development of economic. The famous examples are the stock crashes in the autumn 1929 and 1987 which caused a violent depression in the United States and some other countries, with the collapse of financial markets and the contraction of production and

employment. To alleviate the down influence of market risks, many regulations and methodologies have been proposed since the mid-1990s. In 1988 Basel accord asked banks to deposit 8% capital of risk assets as risk charge. The goal was to restrict the happening of extremely large losses. But it was found soon that the simple 8% rule ignores the diversification of risk by e.g. holding large portfolios, and limits the trading activity of financial institutions. As a remedy, the amendment to the Basel accord officially allowed financial institutions to use their internal models to measure market risks in 1996. This has prompted researches on measuring risks using quantitative methods.

- Credit risk: Risk that opposite partners may not be able to meet their contractual payment obligations. Credit risk includes default risk, country risk and settlement risk. The default risk is one of the main risk resources in credit markets, by which a bond issuer will default, by failing to repay either principal or interest or both on time, see [Duf99]. The settlement risk happens when an expected settlement amount is not being transferred on time. The netting systems is established to minimize this kind of risk, see [KMR03]. The country risk is caused by political and economic uncertainty in a country. It is measured by assessing, among others, the government policies and regulation, economic growth and social stability, see [EGS86].

- Liquidity risk: A financial risk that due to uncertain liquidity. It mainly arises from the unexpected cash outflows of an institution or the low liquidity markets which the institution trades in, see [Dia91].

- Operational risk: The Basel Committee (2004) defines operational risk as the risk of loss resulting from inadequate or failed internal processes, people and systems, or from external events.

- Legal risk: It is risk from uncertainty due to legal actions or uncertainty in the applicability or interpretation of contracts, laws or regulations.

The first two risk categories can be considered as quantifiable risks, which can be measured and expressed by values. The other risk categories, on the other hand, are qualifiable risks, which are measured by experience and market information. This thesis contributes adaptive methodologies for measuring market risk.

2.2 Risk measures

In this thesis the risk management models are mainly implemented to calculate two risk measures value-at-risk (VaR) and expected shortfall (ES) that are based on the distribution of returns.

Definition 2.1: Value-at-Risk (VaR)
Given some probability level $\text{pr} \in [0, 1]$, the VaR of the concerned portfolio is the upper-bound u of the portfolio returns such that the probability of returns r smaller than u is not larger than pr:

$$\text{VaR}_{\text{pr}} = -\inf\{u \in \mathbb{R}, P(r \leq u) \leq \text{pr}\} \tag{2.1}$$

This definition of VaR coincides with the definition of pr-quantile given the distribution of return $r(t)$. Therefore, the VaR can be also defined as:

$$\text{VaR}_{t,\text{pr}} = -\text{quantile}_{\text{pr}}\{r(t)\} \tag{2.2}$$

where pr is the forecasted probability of the portfolio returns over a target time horizon, e.g. $h = 1$ day. Regulator and supervisor consider various probabilities and time horizons according to different risk controlling requirements. Since 1996 banks that are subject to a credit risk charge are allowed to use an "internal model" to calculate the VaRs over time. In Germany for example, the "Grundsatz I" formulated by Bundesaufsichtsamt für Kreditwesen in year 2000 verifies an internal model by setting $h = 1$ days horizon and $\text{pr} = 1\%$ probability level. Internal supervisors of financial institutions with high credit rating such as AA or AAA (Standard & Poors) normally concern very extreme situations with $\text{pr} = 0.5\%$ probability and $h = 1$ days horizon.

It has been known that VaR only tells us the minimal loss in the prefixed "bad" cases, i.e. at pr probability, it however informs less about the size of loss. In other words, VaR is not appropriate for the measurement of capital adequacy and used in the context of portfolio optimization or hedging, see [Jas01]. For this reason, another risk measure, the expected shortfall (ES) was introduced to inform the size of loss.

Definition 2.2: Expected shortfall (ES)
Given the value of VaR at the probability level $\text{pr} \in [0, 1]$ over a target time horizon, the ES is the expected value of the losses (negative returns) exceeding the VaR:

$$\text{ES} = \mathsf{E}\{-r(t) | -r(t) > \text{VaR}_{t,\text{pr}}\} \tag{2.3}$$

Given the definitions of the two popular risk measures, it is clear that both are related to the density of risk factors, e.g. the returns. The relation of these two risk measures is demonstrated in Figure 2.1, by which

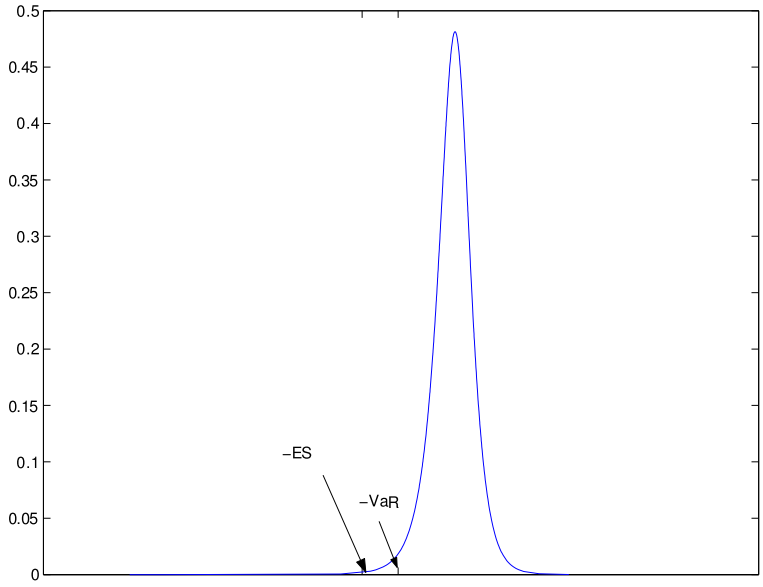

Figure 2.1: Empirical density of the German stock Allianz from 1988/01/04 to 1996/12/30. The values of VaR (average value: 0.035) and ES (0.044) correspond to a probability level pr = 1% over a target time horizon $h = 1$.

the empirical density $\hat{f}_x$ of the daily German stock Allianz from 1988/01/04 to 1996/12/30 is estimated by the nonparametric kernel density estimation: $\hat{f}_x = \frac{1}{Th'} \sum_{t=1}^{T} K\left(\frac{x-x(t)}{h'}\right)$ with bandwidth $h' = 0.661$ and the Quartic kernel function $K(\cdot)$. We refer to [HMSW04] for the choices of bandwidth and kernel function. Given the probability level pr = 1% with the target horizon $h = 1$, the average value of the VaR is 0.035 over the nine years and the ES is 0.044, which are labelled in the figure. The average VaR is smaller than the ES, the expected size of loss. This observation supports the argument before that VaR-based capital reserve underestimates the size of loss and may cause a wrong decision on the risk management. It therefore recommends to consider these two risk measures in risk analysis.

2.3 Requirements on risk analysis

Risks are measured and controlled for regulatory, internal supervisory and making investment decision. Here we discuss the requirements of the three purposes.

Regulatory requirement: As mentioned before, the main goals of risk

No. exceptions	Increase of M_f	Zone
0 bis 4	0	**green**
5	0.4	**yellow**
6	0.5	**yellow**
7	0.65	**yellow**
8	0.75	**yellow**
9	0.85	**yellow**
More than 9	1	**red**

Table 2.1: Traffic light as a factor of the exceeding amount, cited from Franke, Härdle and Hafner (2004).

regulatory are to ensure the adequacy of capital and restrict the happening of large losses of financial institutions. For this reason, it requires the institutions to control their risks at a prescribed level pr and reserve appropriate amount of capital. According to the modification of the Basel market risk paper in 1997, internal models for risk management are verified in accordance with the "traffic light" rule. This rule counts the number of exceptions over VaR at 1% probability spanning the last 250 days and classifies the concerned quantitative model to one colored zone, i.e. green for verified model, yellow for middle-class model and red for problematic model. More important, the rule identifies the multiplicative factor M_f in the market risk charge calculation:

$$\text{Risk charge}(t) = \max\left(M_f \frac{1}{60}\sum_{i=1}^{60} \text{VaR}_{t-i}, \text{VaR}_t\right) \tag{2.4}$$

The multiplicative factor M_f has a floor value 3. It increases corresponding to the number of exceptions, see Table 2.1. For example, if an internal model generates 7 exceptions at 1% probability over the last 250 days, the model is in the yellow zone and its multiplicative factor is $M_f = 3.65$. Financial institutions, whose internal models are located in the yellow or red zone, are normally required to reserve more risk capital than their internal-model-based VaRs with a high probability. Notice that the increase of risk charge will reduce the ratio of profit since the reserved capital can not be invested. On the meanwhile, an internal model is automatically accepted if the number of exceptions does not exceed 4. Remember that the smaller the probability is, the larger is the corresponding VaR, i.e. the absolute value of the quantile. A large VaR requires a large amount of capital reserved and results in a low ratio of profit. In this sense, this regulatory rule in fact suggests banks to

control VaR at 1.6% ($\frac{4}{250}$) probability and reserve risk charge based on the value. Therefore an internal model is particularly desirable by fulfilling the minimum regulatory requirement, i.e. with an empirical probability that is smaller or equal to 1.6%, and simultaneously requiring risk charge as small as possible.

Internal supervisory: It is first important for internal supervisory to exactly measure the market risk exposures before controlling them. In practice, market risk of some kinds of portfolios can be reduced by frequent position adjustments. For example, risks of option portfolio eliminated by hedging delta, gamma and vega, see [Hul97]. However transaction costs make the continuous rebalancing very expensive. Traders tend to use quantitative model to measure the risks of the holding portfolios. If the risks are acceptable, no adjustments are made to the portfolio. Otherwise, traders will reallocate the positions. Consequently an internal model that can exactly generate the empirical probability $\hat{\text{pr}}$ as same as the prefixed value is desired:

$$\hat{\text{pr}} = \frac{\text{No. exceedances}}{\text{No. total observations}}$$

Second, it is important to measure the size of loss for internal supervisory. As discussed before, VaR is inappropriate for the measurement of capital adequacy, since it controls only the probability of default, i.e. the frequency of losses. The ES, on the other hand, considers the average losses in the case of default. Therefore given two models with the same empirical probability, the model has a smaller value of ES is considered better than the other.

Investor: Investors suffer loss once bankruptcy happens. Even in the best case, their loss equals to the difference between the total loss and the reserved risk capital, i.e. the value of ES. Generally risk-averse investors care the amount of loss and thus prefer an internal model with small value of ES. Risk-seeking investors, on the other hand, care profits and hence the small value of risk charge favors their requirements.

2.4 Review of popular risk management models: pros and cons

Prompted by the requirements of regulatory, internal supervisory and investors, many articles contributed methodologies in the popular journal for risk management. In the following, we present several popular risk management methods and discuss their desirable features and disadvantages.

2.4.1 Univariate risk management model

The popular models consider either univariate or multivariate time series. Since the return of portfolio at time point t can be considered as a scalar and its density may be estimated by constructing hypothetical returns given the current trading strategy. One can use the simplified calculation to avoid the covariance estimation and joint density identification. This is called "historical simulation" method. In the univariate space, it is easy to apply complicated but accurate volatility estimation methods and distributional assumptions.

Historical simulation method: given trading strategy $b(t) = \{b_1(t), \cdots, b_d(t)\}^\top$, the d-dimensional risk factors are considered as a univariate time series:

$$r(t) = b(t)^\top x(t) = \sigma_r(t)\varepsilon_r(t) \tag{2.5}$$

by which $\sigma_r(t)$ denotes the volatility of the underlying series $r(t)$ and $\varepsilon_r(t)$ specifies the stochastic feature of the portfolio returns. In the historical simulation it needs to construct the hypothetical portfolio returns with the time-dependent trading strategy $b(t)$ at every time point t. These hypothetical historical observations are assumed to be i.i.d. and follow (2.5).

This method simplifies the calculation and gives a good overview on the risk exposure of the holding portfolio. However, it first requires a large data bank to construct the historical portfolio returns. This requirement is hard to fulfill for over-the-counter (OTC) trading and new markets such as energy market. Furthermore, this method ignores the covariance of risk factors and often results in a low accuracy of estimation.

Historical simulation methods are various due to volatility estimation methods and distributional assumptions. Among others, the moving window average (MWA) and exponential smoothing (ES) methods are the most popular tools for the volatility estimation:

$$\begin{aligned} \text{MWA:} \quad \sigma_r^2(t) &= \frac{1}{M}\sum_{m=0}^{M} r^2(t-m-1), \quad M \le t-2 \\ \text{ES:} \quad \sigma_r^2(t) &= \sum_{m=0}^{\infty} \eta^m r^2(t-m-1) / \sum_{m=0}^{\infty} \eta^m, \quad \eta \in [0,1] \end{aligned}$$

To implement the estimation methods, one needs to specify the value of M in the MWA method or η in the ES method. Several rule-of-thumb values have been proposed such as $M = 250$ in the backtesting process, see the Grundsatz I (Bundesaufsichtsamt für Kreditwesen 2000), and $\eta = 0.94$ for a horizon of one day and $\eta = 0.97$ for a horizon of one month suggested by JP

Morgan, see [Wil00]. As discussed in the Chapter 1, these estimations limit in the fixed form. The accuracy and sensibility of estimation are both lower than those with locally adaptive parameters. We detail the local parameter choice in the later chapters.

Furthermore, the portfolio returns are identified in various distributional framework. The Gaussian distribution is widely used due to its well-known statistical properties. Unfortunately, the risk management models based on this distribution family often underestimate risks since financial time series have heavier tails relative to the Gaussian random variable. Therefore, the Student-t distribution with degrees of freedom 6 and the generalized hyperbolic (GH) distribution have been considered and compared to the Gaussian one. In the thesis, we consider the following historical simulation models and will implement them in the later real data analysis:

- Moving window average with Gaussian, Student-$t(6)$ and GH distributional assumption, denoted as **HS-MAN, HS-MA**$t(6)$ **and HS-MAGH**.
- Exponential smoothing with Gaussian, Student-$t(6)$ and GH distributional assumption, denoted as **HS-RM, HS-ES**$t(6)$ **and HS-ESGH**.

2.4.2 Multivariate risk management models

Multivariate risk management models are often applied by traders, since these multivariate models give more accurate results than the univariate one by considering the correlation of components. On the other hand, these models are mainly used for low-dimensional risk analysis due to the numerical difficulty in the calculation. Recall that the heteroscedastic model w.r.t. the d-dimensional risk factors $x(t)$:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t)$$

where $\Sigma_x(t)$ is the covariance matrix of the risk factors at time point t and $\varepsilon_x(t)$ is an innovation vector with d components.

There are two kinds of methods, i.e. the Gaussian-based methods and the simulation-based methods, used to measure risk of high-dimensional portfolios. For numerical simplification, the risk factors are often assumed to be Gaussian distributed, where given the estimate $\hat{\Sigma}_x(t)$ of the covariance $\Sigma_x(t)$, the standardized returns $\hat{\varepsilon}_x(t) = \hat{\Sigma}_x^{(-1/2)}(t)x(t)$ are independent in the Gaussian distributional framework. As discussed before, the Gaussian assumption is unrealistic. In practice, different distribution family such as the Student-t distribution has been used to identify the marginal distributional behavior

and approximate the distributional behavior of the portfolio in the Monte Carlo simulation method. Similarly, one can use the bootstrap method, by which the samples are repeatedly and randomly generated from the historical data. The limitations of these methods are either due to the unrealistic distributional assumption or cumbersome computation or both.

Gaussian-based method: According to the one-dimensional Taylor expansion, the value of portfolio can be formulated by its derivatives. These derivatives are called "Greeks" in financial study. One famous Gaussian-based risk management model is the Delta-Gamma-Normal model:

$$v(t) = C + \Delta^\top x(t) + \frac{1}{2} x^\top(t) \Gamma x(t)$$

where $v(t)$ is the value function of the portfolio. The first derivative is called delta: $\Delta = \frac{\partial p}{\partial s}$, the rate of the price change of the portfolio w.r.t. the price of the financial instruments. The delta has a value of 1 by holding stock portfolio. The second derivative is the gamma: $\Gamma = \frac{\partial \Delta}{\partial s}$, the rate of the delta change w.r.t. the price of the underlying assets. Given a stock portfolio, the second derivative is 0. Equivalent to say, the Delta-Gamma-Normal model w.r.t. stock portfolios is $v(t) = x(t)$ without loss of generality. It can be once again analyzed in the heteroscedastic model $x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t)$ with $\varepsilon_x(t) \sim \mathrm{N}(0, I_d)$. Now the model only relies on the covariance estimation. Among many others, the RiskMetrics produced by JP Morgan in 1994 has been widely adopted for use on trading floors. In the RiskMetrics, the volatility is first estimated by using the exponential smoothing with a fixed smoothing parameter, e.g. $\eta = 0.94$ for daily returns. The correlation is estimated for $i, j = 1, \cdots, d$ and $i \neq j$ as:

$$\sigma^2_{x_i,x_j}(t) = \{\sum_{m=0}^{M_2} \eta^m x_i(t-m-1)x_j(t-m-1)\} / \{\sum_{m=0}^{M_2} \eta^m\},$$

where the smoothing window is truncated at M_2 such that $\eta^{(M_2+1)} \leq c \to 0$. There is however no guarantee that the estimated covariance is a positive-definite matrix. The dynamic conditional correlation (DCC) model proposed by [ES01] is successful in dealing with this problem, where the covariance matrix is approximated by product of a diagonal matrix and a correlation matrix: $\Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top$, with the diagonal matrix $D_x(t)$ and the correlation matrix $R_x(t)$. Therefore, we consider using the DCC to estimate covariance in this thesis. The method is abbreviated as **DCCN**.

Monte Carlo (MC) simulation method: This method is distributional-free and widely used in practice, by which the covariance $\Sigma_x(t)$ is estimated by e.g. the DCC method, and the distributional behaviors of the standardized returns $\hat{\Sigma}_x^{-1/2}(t)x(t)$ are identified in a prescribed stochastic framework.

Stochastic innovations with S observations are intensively simulated N times based on the identified distribution. The risk measures of portfolio returns at pr probability are calculated empirically as:

$$\begin{aligned} r^{(n)}(t) &= b(t)^\top x^{(n)}(t) = b(t)^\top \hat{\Sigma}_x^{(1/2)}(t)\varepsilon_x^{(n)}(t) \\ VaR_{\mathrm{pr},t} &= -\frac{1}{N}\sum_{n=1}^{N} \mathrm{quantile}_{\mathrm{pr}}\{r^{(n)}(t)\}. \end{aligned}$$

2.5 Adaptive risk management models

The primary aim of this thesis is to present accurate and fast risk management models, which alleviate the limitations of the popular models by following a different thought of research. To be more specific, the basic idea is that risk factors $x(t) \in \mathbb{R}^d$ are represented by a linear combination of d-dimensional independent components (ICs). The linear transformation matrix W is assumed to be nonsingular and estimated by using the independent component analysis (ICA). Due to the independence property, the covariance $D_y(t)$ of the ICs $y(t)$ is a diagonal matrix and the elements of the stochastic vector $\varepsilon_y(t)$ are cross independent. From a statistical viewpoint, this projection technique is desirable since the d-dimensional portfolio is decomposed to univariate and independent risk factors through a simple linear transformation. The joint density (f) and the covariance of any linear transformed ICs such as $x(t) = W^{-1}y(t)$ are analytically computable:

$$\begin{aligned} f_y &= \prod_{j=1}^{d} f_{y_j}, & \Sigma_y(t) &= D_y(t) \\ f_x &= \mathrm{abs}(|W|) f_y(Wx), & \Sigma_x(t) &= W^{-1} D_y(t) W^{-1\top}. \end{aligned}$$

Furthermore the diagonal elements of $D_y(t)$ and each component of $\varepsilon_y(t)$ are estimated univariately since the matrix manipulation is equivalent to:

$$y_j(t) = \sigma_{y_j}(t)\varepsilon_{y_j}(t), \quad j = 1, \cdots, d, \tag{2.6}$$

where $\sigma_{y_j}(t)$ is the square root of the j-th diagonal element in $D_y(t)$ and $\varepsilon_{y_j}(t)$ is the univariate stochastic term. In this thesis, two univariate risk management models are considered:

- **Adaptive local constant model with GH distributional assumption (GHADA)** Volatility is estimated by using the local constant method with the time-dependent parameter $M(t)$ and the innovations are assumed to be GH distributed.

- **Local exponential smoothing with GH distributional assumption (LESGH)** Volatility is estimated by using the local exponential smoothing method with the time-dependent parameter $\eta(t)$ and the innovations are assumed to be GH distributed.

After the ICA and the univariate case study on the ICs, we first approximate the risk measures based on the Monte Carlo simulation, see [CHS6a] . In particular, we generate $d-$dimensional samples of the fitted distributions with sample size S, from which we calculate the daily empirical pr-quantile of the portfolio variations. The simulation will repeat N times and the average value of the empirical quantiles is considered as the portfolio VaR at level pr:

$$\hat{\mathrm{VaR}}_{\mathrm{pr},t} = -\frac{1}{N}\sum_{n=1}^{N} \hat{F}^{-1}_{\mathrm{pr},t}\{r^{(n)}(t)\} = \frac{1}{N}\sum_{n=1}^{N} \hat{F}^{-1}_{\mathrm{pr},t}\{b(t)^{\top}\hat{W}^{-1}\hat{D}_y^{1/2}(t)\hat{\varepsilon}_y^{(n)}(t)\}$$

where $\hat{F}^{-1}_{\mathrm{pr},t}$ denotes the empirical quantile function of R_t.

We name this procedure **ICVaR method**. The algorithm is briefly summarized in the following:

1. Implement ICA to get independent components.
2. Apply the GHADA method to estimate the local constant volatility and fit the marginal pdf of $\varepsilon_y(t)$.
3. Determine VaR at level pr via MC simulation. In each scenario, we generate d dimensions series with S observations. The scenarios are repeated N times.
4. Calculate the ES.

Moreover, [CHS6b] introduce another simple and fast multivariate risk management method, by implementing the ICA to the high-dimensional series and fitting the resulting ICs in the GH distributional framework as well. The named **GHICA method** improves the ICVaR method from two aspects. The volatility estimation is driven by the local exponential smoothing technique to achieve the best possible accuracy of estimation. The fast Fourier transformation (FFT) technique is used to approximate the density of the portfolio returns. Compared to the Monte Carlo simulation technique used in the former study, it significantly speeds up the calculation.

The algorithm is summarized in the following:

1. Do ICA to the given risk factors to get ICs.
2. Implement local exponential smoothing to estimate variance of each IC

3. Identify the innovations of each IC in the GH distributional framework
4. Estimate the density of the portfolio return using the FFT technique
5. Calculate risk measures

Part II

Adaptive Risk Management - Univariate Models

Chapter 3

Adaptive risk management 1: GHADA

3.1 Introduction

After the breakdown of the fixed exchange rate system of the Bretton Woods agreement in 1971, a sudden increase of volatility was observed in financial markets. The following boom of financial derivatives accelerated the turbulence of the markets. The subsequent scale of losses astonished the world and pushed the development of sound risk management systems. One of the most challenging tasks in analyzing financial markets is to measure and manage risks properly. Financial risks have many sources and are typically mapped into a stochastic framework:

$$r(t) = \sigma_r(t)\varepsilon_r(t), \tag{3.1}$$

where $r(t)$ denotes the returns of financial instrument: Given the price $s(t)$ of the financial instrument, $r(t) = \log\{s(t)/s(t-1)\}$. The volatility $\sigma_r(t)$ is time-dependent and $\varepsilon_r(t)$ is the white noise. For notational simplification, we eliminate the subscript r in this chapter, by writing $r(t) = \sigma(t)\varepsilon(t)$.

In risk management, various risk measures are calculated based on the estimated distribution of the risk factor $r(t)$. Among others, Value at Risk (VaR) has become the standard measure of market risk since JP Morgan launched RiskMetrics in 1994. For a given financial instrument, VaR indicates the possible loss at a certain risk level over a certain time horizon. VaR at a risk level pr is defined in a mathematical form:

$$\mathrm{VaR}_{t,\mathrm{pr}} = -F^{-1}_{\mathrm{pr},t}\{r(t)\} = -\sigma(t)\mathrm{quantile}_{\mathrm{pr}}\{\varepsilon(t)\}$$

where F_t^{-1} is the quantile function of $r(t)$ at time t, which is equal to the product of the volatility and the pr-th quantile of the stochastic term $\varepsilon(t)$,

[FHH04]. The importance of VaR was reinforced after it was used by central banks to govern and supervise the capital adequacy of banks in the Group of Ten (G10) countries in 1995. However VaR concerns more on the frequency of loss and is inappropriate that measures capital adequacy, see the discussion in Chapter 2. Therefore, risk analysts consider as well the expected shortfall (ES) to measure the expected size of loss in the case that the realized loss exceeds the VaR at the prescribed probability level pr:

$$\mathrm{ES} = \mathsf{E}\{-r(t)| - r(t) > \mathrm{VaR}_{t,\mathrm{pr}}\}$$

Indicated by (3.1), the estimated density, on which these two risk measures rely, inherently depends on the distributional assumption of the stochastic term and the volatility estimation.

In literature, for reasons of stochastic and numerical simplicity, it is often assumed that the involved risk factors are normally distributed e.g. in the RiskMetrics framework. With the distributional assumption of normality, the RiskMetrics makes the analysis of VaR simple and standard, [Jor01]. [ABCD05] have pointed out that returns will converge to normality under temporal aggregation. This observation verifies the principle of the RiskMetrics method when a long time horizon such as two weeks $h = 10$ or one month $h = 25$ is considered. On the other hand, financial institutions concern more on their daily or weekly risk exposures. The daily returns however present distinct behavior from the Gaussian random variables. It is found that at high probability level such as $\mathrm{pr} = 5\%$, the Gaussian-based VaR is very close to that under realistic heavy-tailed distribution, see [JJ02]. In the "worse" case such as $\mathrm{pr} = 1\%$, the discrepancy between the Gaussian-based VaRs and the realized VaRs over time becomes evident. Since more and more extreme losses were observed in the market, risks at low probability have drawn the attention of risk analysts.

This empirical fact is illustrated on the basis of the daily foreign exchange (FX) rates of the German Mark to the US Dollar (DEM/USD) from 1979/12/01 to 1994/04/01. Suppose the volatility estimates $\hat{\sigma}(t)$ are given, the standardized returns $\hat{\varepsilon}(t) = r(t)/\hat{\sigma}_t$ eliminate the influence of volatility clustering and are more stationary than the returns. The technique used to estimate volatility $\sigma(t)$ will be discussed later. The density estimations of the standardized returns are displayed in Figure 3.1. The nonparametric kernel density of the standardized returns is regarded as benchmark, where the Quartic kernel function is used and the Silverman's rule of thumb is applied to choose the bandwidth, [HMSW04]. The log scale of the density is depicted as well. The estimated Gaussian density obviously deviates from the benchmark, which will lead to inaccurate VaR calculations. Due to the weak ability

of Gaussian distributions in capturing this empirical distributional feature of financial risk factors, various heavy-tailed distribution families such as the hyperbolic and Student-t distributions and the Lévy process have been applied in finance by [EK95a], [EMS99] and [BNS01]. Among them, the generalized hyperbolic (GH) distribution family, with five parameters, can well match the distributional behavior of real data in a flexible way, see [EKK03]. The figure shows the case, by which the identified GH distribution coincides the benchmark.

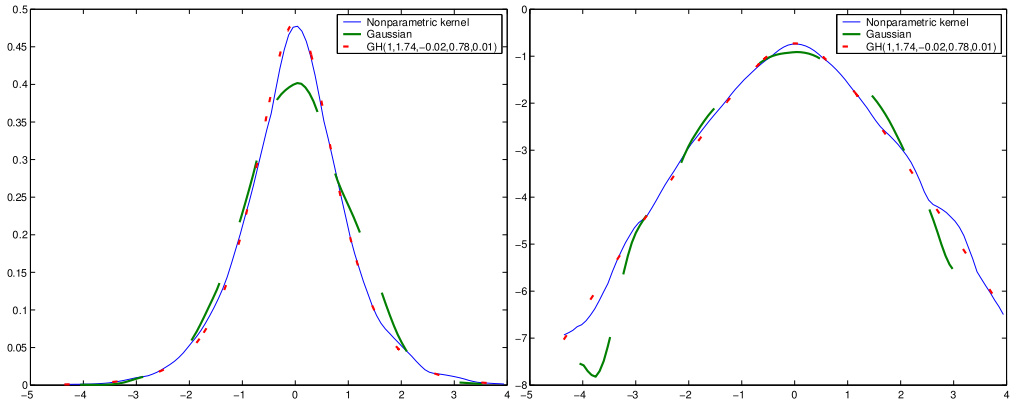

Figure 3.1: Density estimation of the daily DEM/USD standardized returns from 1979/12/01 to 1994/04/01 (3719 observations). The log scale of the estimation is displayed on the right panel. The nonparametric kernel density estimate is considered as benchmark. The bandwidth is $h' \approx 0.54$. Data source: FEDC (http://sfb649.wiwi.hu-berlin.de).

The second important task in risk analysis, as discussed before, is to estimate volatility in a meaningful way. Many methodologies have been proposed in this field. In summary, most of them have two limitations: They either rely on a time-invariant model or are based on an unrealistic distributional assumption. For example, the most frequently used estimations are the ARCH [Eng95], GARCH [Bol95] and stochastic volatility models [HRS95]. These estimations reflect the volatility clustering of financial time series, they are however not flexible enough to react to possible structure shifts of volatility process. It is often questionable to use a time constant closed form in estimation, especially in long time periods.

In the recent, flexible estimation methods have been presented by providing a data-driven "local" model. For example, the moving window average

model	volatility estimation	distributional assumption
HS-RM	GARCH(1,1)	Gaussian
HS-ESt(6)	GARCH(1,1)	Student-t(6)
HS-ESGH	GARCH(1,1)	GH
HS-ADAN	local constant	Gaussian
HS-ADAt(6)	local constant	Student-t(6)

Table 3.1: Alternative risk management models.

method is popular with a simple form:

$$\sigma^2(t) = \frac{1}{M}\left\{\sum_{m=0}^{M} r^2(t-m-1)\right\}, \quad M \leq t-2$$

Given a fixed value of M, this estimation method in fact assumes that the volatility is time homogeneous over the last M time points. It arises a question how to choose the smoothing parameter M. [MS04a] present the local constant model by choosing the smoothing parameter M at every time point t:

$$\sigma^2(t) = \frac{1}{M(t)}\left\{\sum_{m=0}^{M(t)} r^2(t-m-1)\right\}, \quad M(t) \leq t-2$$

It improves the flexibility of the moving window average method and can quickly react to a sudden structure shift. Despite the improved accuracy of estimation, the work is based on the Gaussian distribution, which is unrealistic for financial time series.

Motivated by the above two lines of research, we extend the local volatility estimation in the GH distributional framework. We name this new VaR technique the **G**eneralized **H**yperbolic **Ada**ptive Volatility (GHADA) technique. The standardized return density plot in Figure 3.1 is actually calculated with the GHADA technique. In this chapter, we also compare the risk calculation based on the GHADA technique and on several alternative risk management models. The alternative models are classified according to the volatility estimation and the distributional assumption. For example, the RiskMetrics (RM) implements the GARCH(1,1) to do the volatility estimation. Notice that these methods belong to the historical simulation method, namely analyzing the portfolio returns in univariate space, see Chapter 2 for more details. We name these alternative methods with the abbreviation **HS** in the beginning, see Table 3.1.

The chapter is organized as follows: in Section 3.2, we introduce the details of the GHADA technique. The validation of the GHADA model is

illustrated through simulation study in Section 3.3. In Section 3.4, VaR and ES are calculated based on the DEM/USD and German bank portfolio data. According to the backtesting results, the GHADA technique provides more accurate forecasts than the models with assumptions of the Gaussian and Student-t distributions. Furthermore in extreme events, the GHADA technique performs better than the models with GARCH(1,1) volatility processes. Finally, we conclude our study in Section 3.5.

3.2 GHADA technique

In risk management modelling, a major task is to appropriately forecast the return's distribution. According to the VaR and ES definitions, the distributional assumption of the stochastic term $\varepsilon(t)$ and the volatility estimation influence the accuracy of risk analysis. In this section, we describe two pillars of the proposed GHADA technique: the GH distribution and the local constant volatility estimation.

3.2.1 Generalized hyperbolic distribution

The GH distribution introduced by [BN77] is a heavy-tailed distribution that can well replicate the empirical distribution of financial risk factors. The density of the GH distribution for $x \in \mathbb{R}$ is:

$$f_{GH}(x;\lambda,\alpha,\beta,\delta,\mu) = \frac{(\iota/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\iota)} \frac{K_{\lambda-1/2}\left\{\alpha\sqrt{\delta^2+(x-\mu)^2}\right\}}{\left\{\sqrt{\delta^2+(x-\mu)^2}/\alpha\right\}^{1/2-\lambda}} \cdot e^{\beta(x-\mu)} \quad (3.2)$$

under the conditions:

$$\begin{aligned} &\delta \geq 0, \quad |\beta| < \alpha \quad \text{if } \lambda > 0 \\ &\delta > 0, \quad |\beta| < \alpha \quad \text{if } \lambda = 0 \\ &\delta > 0, \quad |\beta| \leq \alpha \quad \text{if } \lambda < 0 \end{aligned}$$

where λ, α, β, δ and $\mu \in \mathbb{R}$ are the GH parameters with $\iota^2 = \alpha^2 - \beta^2$. The density's location and scale are mainly controlled by μ and δ respectively:

$$\begin{aligned} \mathsf{E}[X] &= \mu + \frac{\delta^2\beta}{\delta\iota}\frac{K_{\lambda+1}(\delta\iota)}{K_{\lambda}(\delta\iota)} \\ \mathrm{Var}[X] &= \delta^2\left\{\frac{K_{\lambda+1}(\delta\iota)}{\delta\iota K_{\lambda}(\delta\iota)} + (\frac{\beta}{\iota})^2[\frac{K_{\lambda+2}(\delta\iota)}{K_{\lambda}(\delta\iota)} - \{\frac{K_{\lambda+1}(\delta\iota)}{K_{\lambda}(\delta\iota)}\}^2]\right\}, \end{aligned}$$

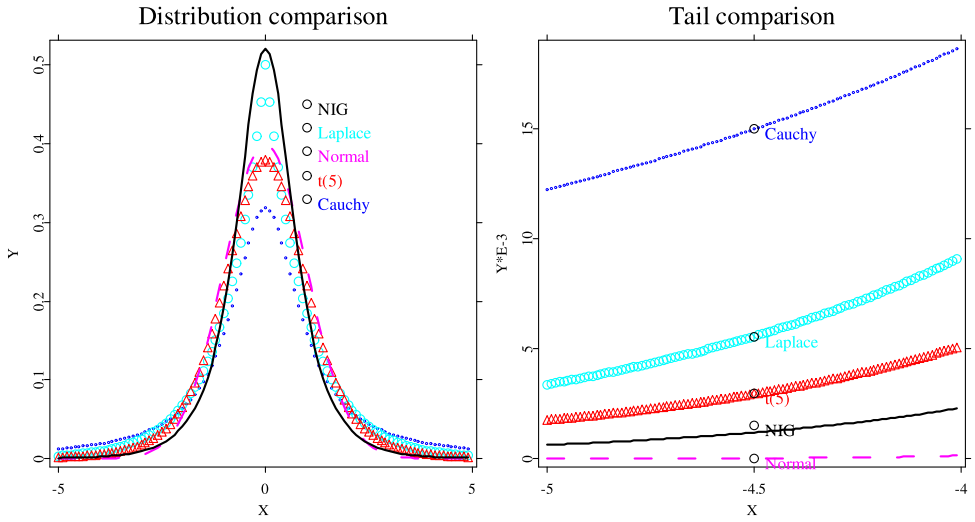

Figure 3.2: Tail-behavior of five standardized distributions: NIG distribution, standard Gaussian distribution, Student-t distribution with degrees of freedom 5, Laplace distribution and Cauchy distribution.

whereas β and α play roles in the skewness and kurtosis of the distribution. For more details of the parameters' domains, we refer to [BS01]. $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ, [BNB81]:

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} exp\{-\frac{x}{2}(y+y^{-1})\}\, dy$$

Furthermore, the GH distribution has a tail behavior:

$$f_{GH}(x;\lambda,\alpha,\beta,\delta,\mu=0) \sim x^{\lambda-1}e^{(\mp\alpha+\beta)x} \text{ as } x \to \pm\infty, \tag{3.3}$$

where $a(x) \sim b(x)$ as $x \to \infty$ meaning that both $a(x)/b(x)$ and $b(x)/a(x)$ are bounded as $x \to \infty$. Recall that the large losses locate in the left tail of the return's distribution. The right tail, on the other hand, concerns the values of profits that are less interesting in risk management.

We compare the GH distribution with other three popular heavy-tailed distributions, the Student-t, Laplace and Cauchy distributions, and the Gaussian distribution. In order to guarantee the comparability of these distributions, we concern the standardized variables with mean of 0 and variance of 1. Here one important subclass of the GH distribution is considered, i.e. the normal inverse Gaussian (NIG) distribution with $\lambda = -\frac{1}{2}$, which is introduced more precisely in the following text. In the left panel of Figure 3.2, the complete shapes of these distributions are displayed. Among them,

the Cauchy distribution has the lowest peak and the fattest tails. Briefly to say, it has the flattest distribution. The NIG density displays a faster speed of exponential decay than other three heavy-tailed distributions and has the highest peak, which is similar to the distributional shape of real data, see e.g. Figure 3.1. It states that the GH distribution better matches the empirical tail behavior of real data than the alternatives.

The moment generating function of the GH distribution is:

$$m_f(z) = e^{\mu z} \cdot \frac{\iota^\lambda}{\iota_z^\lambda} \cdot \frac{K_\lambda(\delta \iota_z)}{K_\lambda(\delta \iota)}, \quad |\beta + z| < \alpha,\ \iota_z^2 = \alpha^2 - (\beta + z)^2 \tag{3.4}$$

indicating that m_f is differentiable infinitely many times near 0. As a result, every moment of a GH variable exists. In Section 3.2.2, this feature as well as the tail behavior in (3.3) of the GH distribution helps to extend the local constant volatility methodology from the Gaussian distribution to the GH distribution, see Appendix 3.6.

Given the closed form of the GH distribution (3.2), it is tractable to apply the maximum likelihood (ML) estimation to identify the distributional parameters. Such a direct estimation based on the GH distribution is however difficult since the estimation of λ in the modified Bessel function is computationally cumbersome and numerically unstable. Instead, subclasses of the GH distribution such as the hyperbolic (HYP) and Gaussian-inverse Gaussian (NIG) distributions are frequently used. These subclasses prefix the value of λ to avoid the numerical problem. [EK95a] and [BN97] have shown that these subclasses are rich enough to model financial time series in an efficient way. In addition, the popularity of the subclasses is also motivated by the observation that the four parameters $(\mu,\, \delta,\, \beta,\, \alpha)^\top$ can simultaneously control the four moment functions of the distribution, i.e. the trend, scale, asymmetry and likeliness of extreme events. In our study we concentrate on two subclasses of the GH distribution: HYP with $\lambda = 1$ and NIG distribution with $\lambda = -1/2$. The corresponding density functions are given as:

- Hyperbolic (HYP) distribution: $\lambda = 1$,

$$f_{HYP}(x; \alpha, \beta, \delta, \mu) = \frac{\iota}{2\alpha\delta K_1(\delta\iota)} e^{\{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\}}, \tag{3.5}$$

  where $x, \mu \in \mathbb{R}$, $0 \le \delta$ and $|\beta| < \alpha$,

- Normal-inverse Gaussian (NIG) distribution: $\lambda = -1/2$,

$$f_{NIG}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} \frac{K_1\left\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\right\}}{\sqrt{\delta^2 + (x-\mu)^2}} e^{\{\delta\iota + \beta(x-\mu)\}}. \tag{3.6}$$

  where $x, \mu \in \mathbb{R}$, $\delta > 0$ and $|\beta| \le \alpha$.

In order to estimate the unknown parameters $(\alpha, \beta, \delta, \mu)^\top$, ML estimation and numerical optimization methods such as the Powell method [PTVF92] are used. For an independently and identically distributed (i.i.d.) HYP (3.7) and NIG (3.8) distributed variable, the log-likelihood functions are:

$$\begin{aligned} L_{HYP} &= T\log\iota - T\log 2 - T\log\alpha - T\log\delta - T\log K_1(\delta\iota) \\ &\quad + \sum_{t=1}^{T}\left[-\alpha\sqrt{\delta^2 + \{x(t)-\mu\}^2} + \beta\{x(t)-\mu\}\right] \end{aligned} \tag{3.7}$$

$$\begin{aligned} L_{NIG} &= T\log\alpha + T\log\delta - T\log\pi + T\delta\iota \\ &\quad + \sum_{t=1}^{T}\log K_1\left\{\alpha\sqrt{\delta^2 + \{x(t)-\mu\}^2}\right\} \\ &\quad \sum_{t=1}^{T}\left[-\frac{1}{2}\log\left[\delta^2 + \{x(t)-\mu\}^2\right] + \beta\{x(t)-\mu\}\right] \end{aligned} \tag{3.8}$$

Figure 3.1 displays the estimated HYP density with the ML estimators $\hat{\alpha} = 1.744$, $\hat{\beta} = -0.017$, $\hat{\delta} = 0.782$, $\hat{\mu} = 0.012$ of the standardized DEM/USD returns. The estimated density graphically coincides with the empirical density, namely the nonparametric kernel density estimation. The estimated NIG density has the similar behavior as the empirical one, which is eliminated here. Normally, the HYP assumption shows a bit better performance than the NIG, as illustrated in the later simulation and empirical studies. But no evidence shows that one is superior to another.

3.2.2 Adaptive volatility estimation

Next we describe the adaptive estimation procedure for the volatility coefficients when risk factors are GH distributed. The concept of the local constant model is in fact localization of the moving window average. In other words, there exists an interval of local homogeneity of the volatility process $\sigma(t)$: $I = [t - M(t), t)$ such that $\sigma(t)$ varies little over I. Once an interval of homogeneity I, or saying, the value of $M(t)$ is specified, the local volatility is estimated by averaging the squared returns over the time interval I:

$$\hat{\sigma}^2(t) = \frac{1}{M(t)}\sum_{m=0}^{M(t)} r^2(t-m-1). \tag{3.9}$$

The squared returns are always nonnegative and have a skewed distribution with the stochastic errors $\varepsilon(t)$. Therefore, the problem of estimating $\sigma(t)$ is transformed into an additive regression problem by a power transformation:

$$\begin{array}{rcccc} |r(t)|^\gamma & = & C_\gamma\sigma(t)^\gamma & + & D_\gamma\sigma(t)^\gamma\zeta_\gamma(t) \\ & = & \theta(t) & + & s_\gamma\theta(t)\zeta_\gamma(t) \end{array} \tag{3.10}$$

with the power transformation parameter γ. It states in the proof that γ is a constant bounded by $[0, 1)$, see Appendix 3.6. The stochastic term $\zeta_\gamma(t) = (|\varepsilon(t)|^\gamma - C_\gamma)/D_\gamma$. Equation (3.10) can be considered as a regression model to estimate $\theta(t)$ with heteroscedastic additive errors $s_\gamma\theta(t)\zeta_\gamma(t)$. Given a fixed value γ, $C_\gamma = \mathsf{E}(|\varepsilon(t)|^\gamma|\mathcal{F}_{t-1})$, $D_\gamma^2 = \mathsf{E}[(|\varepsilon(t)|^\gamma - C_\gamma)^2|\mathcal{F}_{t-1}]$ and $s_\gamma = D_\gamma/C_\gamma$ are all constants and merely depends on γ. It indicates that the local volatility $\sigma(t)$ has the same time homogeneous interval as $\theta(t)$. Moreover, the resulting $\theta(t)$ can be estimated by the power transformed returns since the stochastic term $\zeta_\gamma(t)$ is martingale difference with expectation of 0. Notice that in the estimation of $\theta(t)$. These constants in the power transformation such as C_γ are not necessarily required for specifying interval of homogeneity.

If I is an interval of homogeneity, then $\theta(t)$ is approximated by a constant denoted by θ_I:

$$\hat{\theta}_I = \frac{1}{M(t)} \sum_{m=0}^{M(t)} |r(t-m-1)|^\gamma, \quad M(t) \leq t-2. \tag{3.11}$$

By (3.10), we have

$$\hat{\theta}_I = \frac{1}{M(t)} \sum_{m=0}^{M(t)} \theta(t-m-1) + \frac{s_\gamma}{M(t)} \sum_{m=0}^{M(t)} \theta(t-m-1)\zeta_\gamma(t).$$

The conditional expectation and variance of $\hat{\theta}_I$ are as follows:

$$\begin{aligned} \mathsf{E}[\hat{\theta}_I|\mathcal{F}_{t-1}] &= \mathsf{E}\,\frac{1}{M(t)} \sum_{m=0}^{M(t)} \theta(t-m-1), \\ v_I^2 = \mathrm{Var}[\hat{\theta}_I|\mathcal{F}_{t-1}] &= \frac{s_\gamma^2}{M^2(t)}\,\mathsf{E}\left\{\sum_{m=0}^{M(t)} \theta(t-m-1)\zeta_\gamma(t)\right\}^2 \end{aligned}$$

Inside the homogeneous interval I, v_I can be estimated by:

$$\hat{v}_I = s_\gamma \hat{\theta}_I M^{-1/2}(t).$$

The homogeneity test of $\theta(t)$ is based on a martingale deviation probability bound with $\varepsilon(t) \sim$ GH. Details are given in the Appendix 3.6. We precisely address the probability of the estimation bias in Theorem 1 below, the proof is similar as in [MS04a].

THEOREM 1 *If the volatility coefficient $\sigma(t)$ satisfies the condition $b \leq \sigma(t)^2 \leq bB$ with some positive constants b and B, then it holds that:*

$$\begin{aligned} P\left(|\hat{\theta}_I - \theta(t)| > \Delta_I\{1 + \mathfrak{z}s_\gamma M^{-1/2}(t)\} + \mathfrak{z}\hat{v}_I\right) \\ \leq 4\sqrt{e}\mathfrak{z}(1 + logB)\exp\left\{-\frac{\mathfrak{z}^2}{2a_\gamma(1 + \mathfrak{z}s_\gamma|I|^{-1/2})^2}\right\} \end{aligned}$$

where Δ_I *is the squared bias defined as* $\Delta_I^2 = \frac{1}{M(t)} \sum_{m=1}^{M(t)} \{\theta(t-m) - \theta(t)\}^2$.

This theorem indicates that, if I is a time homogeneous interval, the squared bias Δ_I is negligible and the estimation error $|\hat{\theta}_I - \theta(t)|$ is small relative to $(\mathfrak{z}\hat{v}_I)$ for $t \in I$ with a high probability. In practice, the candidate homogeneous interval is divided to two subintervals: J and its complement set $I \backslash J$. For any subinterval J of I, it holds with high probability if $\mathfrak{z}' = \mathfrak{z} s_\gamma$ is large enough:

$$\left| \hat{\theta}_{I\backslash J} - \hat{\theta}_J \right| \leq \mathfrak{z} \left(\hat{v}_J + \hat{v}_{I\backslash J} \right) = \mathfrak{z}' \left(\hat{\theta}_J |M_J|^{-1/2} + \hat{\theta}_{I\backslash J} |M_{I\backslash J}|^{-1/2} \right) \tag{3.12}$$

where M_J and $M_{I\backslash J}$ denote the smoothing parameter $M(t)$ in the estimation of $\theta(t)$ w.r.t. the subintervals J and its complement set. Consequently, if there exists a subinterval $J \subset I$ by which $\left| \hat{\theta}_{I\backslash J} - \hat{\theta}_J \right|$ is significantly larger than the right hand side of (3.12), the homogeneity of the interval I is rejected.

There are two parameters involved in the homogeneity test: γ in the power transformation and the critical value $\mathfrak{z}'$ in (3.12). According to Lemma 1, the parameter γ should be bounded by $[0, 1)$. In our study, we choose $\gamma = 0.5$. [MS04a] show that the choice of γ does not have much effect on the procedure for estimating the interval of homogeneity. The critical value $\mathfrak{z}'$ is chosen based on an empirical way, namely minimizing the forecasting errors. In the process, we take t_0 such that there are enough historical observations to estimate the former local volatilities. The value of $\mathfrak{z}'$ that gives minimal forecasting errors is chosen:

$$\mathfrak{z}' = \text{argmin} \sum_{t=t_0}^{T-1} \left\{ |r(t)|^\gamma - \hat{\theta}_{\mathfrak{z}'}(t) \right\}^2 \tag{3.13}$$

where $\hat{\theta}_{(t,\mathfrak{z}')}$ is the estimate of $\theta(t)$ (3.11), which depends on the critical value $\mathfrak{z}'$. This parameter choice is distributional free.

Now we are ready to describe the precise procedure for estimating the interval of homogeneity and further the local volatility, given the estimated critical value $\hat{\mathfrak{z}'}$. We set $k = 1$ and start with a small interval $I = [t - m_0, t)$ that automatically satisfies the homogeneity, e.g. $m_0 = 5$. Stepwise, we increase value of k by 1 and test the homogeneity in a larger interval $I = [t - k \times m_0, t)$. The choice of the step-increasing parameter m_0 influences the sensitivity of the estimation to a change point, for instance, a smaller value increases the sensitivity but slows the estimation speed. This property will be analyzed in Section 3.3. The algorithm is presented in the following. For every time point t,

Step 1 Increase k to $k+1$, and enlarge the interval I to $[t-m, t)$ with $m = k \times m_0$.

Step 2 Pointwise test the homogeneity hypothesis for $J(\ell) = [\tau - \frac{2m}{3} + \ell, \tau)$, $\ell = 1, 2, \cdots, \frac{m}{3}$ and reject the hypothesis if it holds true:

$$\left| \hat{\theta}_{I \setminus J(\ell)} - \hat{\theta}_{J(\ell)} \right| > \hat{\mathfrak{z}}' \left(\hat{\theta}_{J(\ell)} |J(\ell)|^{-1/2} + \hat{\theta}_{I \setminus J(\ell)} |I \setminus J(\ell)|^{-1/2} \right) \tag{3.14}$$

Step 3 If the homogeneity is rejected, the procedure terminates with $M(t) = (k-1) \times m_0$. Otherwise, go back to **Step 1**.

Step 4 Use the smoothing parameter $\hat{M}(t)$ identified in the volatility estimation:

$$\hat{\sigma}^2(t) = \frac{1}{\hat{M}(t)} \sum_{m=0}^{\hat{M}(t)} r^2(t-m-1).$$

3.3 Simulation study

As discussed before, quality of risk management models relies on two factors: volatility estimation and distributional identification of risk factors. The previous distributional estimation based on the DEM/USD data in Section 3.2.1, provides evidence that with four parameters, the HYP and NIG distributions can represent the empirical distribution of the stochastic term very well. In this section, we illustrate the reliability of the local constant volatility model with four different distributional assumptions. Two simple volatility processes with jumps and a GARCH(1,1) process are considered here:

$$\sigma_1(t) = \left\{ \begin{array}{lll} |0.02t - 5|/100 & , & 1 \le t \le 300 \\ |0.02t - 20|/100 & , & 300 < t \le 600 \\ |0.12t - 30|/100 & , & 600 < t \le 1000 \end{array} \right\} \tag{3.15}$$

$$\sigma_2(t) = \left\{ \begin{array}{lll} 0.01 & , & 1 \le t \le 400 \\ 0.03 & , & 400 < t \le 750 \\ 0.015 & , & 750 < t \le 1000 \end{array} \right\} \tag{3.16}$$

$$\sigma_3(t) = 1.65e-06 + 0.07\varepsilon^2(t-1) + 0.89\sigma^2(t-1) \tag{3.17}$$

where the parameters of the GARCH(1,1) process (3.17) are the estimates of the DEM/USD returns from 1979/12/01 to 1994/04/01.

In each scenario, we generate 1000 innovations respectively with assumptions: $\varepsilon(t) \sim \text{HYP}(2,0,1,0)$, $\varepsilon(t) \sim \text{NIG}(2,0,1,0)$, $\varepsilon(t) \sim \text{N}(0,1)$ and

$\varepsilon(t) \sim t(6)$. The risk factors are generated based on the heteroscedastic model:

$$r_{ij}(t) = \sigma_i(t)\varepsilon_j(t), \quad i = 1, 2, 3 \text{ and } j = \text{HYP}, \text{NIG}, \text{N}, t(6).$$

In the local constant (LC) model, the first 200 observations of $r_{ij}(t)$ are considered as a training set. The transformation parameter γ is fixed at 0.5 and a global λ' that minimizes the forecasting error for $t \in [201, 1000]$ is selected for the homogeneity test. Last but not least, two values of the step-increasing parameter m_0 in the homogeneous interval are used since the value of m_0 will influence the detection speed of the LC model for jumps: The recommended value $m_0 = 5$ and a more sensitive value $m_0 = 2$. Note that a smaller value of m_0 in general can find jumps faster than a larger one. All scenarios are repeated $N = 200$ times.

Three examples of the estimated volatility series: $\hat{\sigma}_1(t)$, $\hat{\sigma}_2(t)$ and $\hat{\sigma}_3(t)$ with the HYP or NIG distributional assumption are displayed in Figure 3.3, where the LC and GARCH(1,1) estimations are compared with the true volatilities. In the first two estimations, the LC and GARCH(1,1) models display comparable results, whereas the GARCH(1,1) setup performs better in the estimation of σ_{3t}.

The quality of these two volatility estimation techniques is further measured by two criteria, the ratio of mean absolute error (**RMAE**) and the ratio of mean squared error (**RMSE**):

$$\begin{aligned} \text{RMAE} &= \frac{\sum_{t=201}^{1000} |\hat{\sigma}_i^{LC}(t) - \sigma_i(t)|}{\sum_{t=201}^{1000} |\hat{\sigma}_i^{GARCH}(t) - \sigma_i(t)|}, \\ \text{RMSE} &= \frac{\sum_{t=201}^{1000} (\hat{\sigma}_i^{LC}(t) - \sigma_i)^2(t)}{\sum_{t=201}^{1000} (\hat{\sigma}_i^{GARCH}(t) - \sigma_i)^2(t)}, \quad i = 1, 2, 3 \end{aligned}$$

If the value of RMAE or RMSE is smaller than 1, it means that the LC model has a smaller estimation error on average than the GARCH and vice versa. Based on the 200 repetitions, the mean, standard deviation (sd), maximum and minimum of the two criteria are reported in Table 3.2. In the simulation of σ_{1t}, the LC model with $m_0 = 5$ gives more accurate volatility estimations on average than the GARCH(1,1) technique. Concerning $\sigma_2(t)$, the GARCH model performs better than the LC when the stochastic terms are HYP and NIG distributed, but has lower accuracy when $\varepsilon(t)$ are Gaussian and t distributed. As assumed, the GARCH model can better match the generated GARCH process $\sigma_3(t)$. Based on the simulation results, the LC model is comparable to the GARCH(1,1) technique. However, the assumption of the GARCH technique, i.e. the estimation form is time constant, is disputed as the nonstationary volatility processes $\sigma_1(t)$ and $\sigma_2(t)$ are given. In this

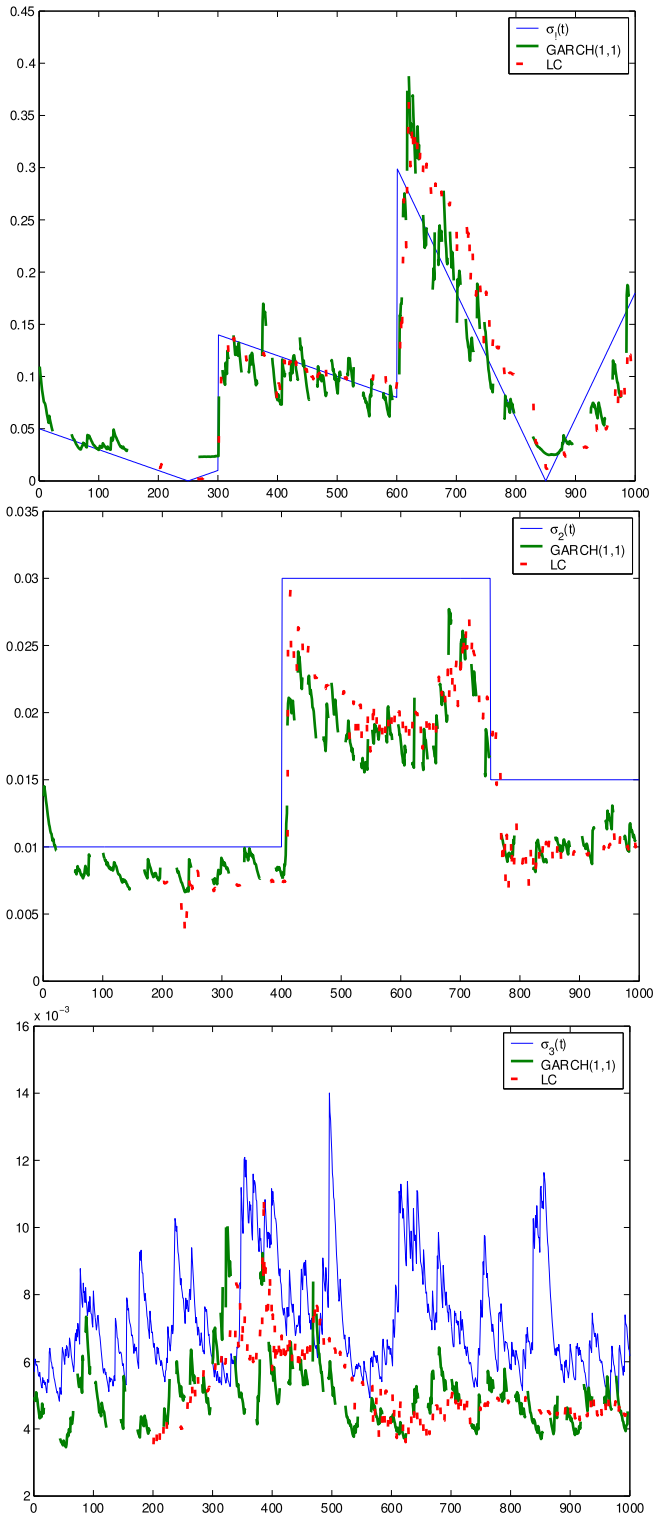

Figure 3.3: One realized estimation based on the simulated data: 1) the HYP variables for $\sigma_1(t)$, 2) the NIG for $\sigma_2(t)$ and 3) the NIG for $\sigma_3(t)$. The involved parameters are $\gamma = 0.5$, $m_0 = 5$ and the starting point $t_0 = 201$. The volatility processes are estimated by using the GARCH(1,1) model and the local constant (LC) model respectively.

sense, the LC model is considered to be better since it not only represents the empirical characteristics of volatility movements, but also provides a reasonable explanation for the estimation.

	RMAE				RMSE			
$\hat{\sigma}_1(t) : \varepsilon(t)$	mean	sd	max	min	mean	sd	max	min
HYP	0.77	1.05	1.04	0.06	0.78	1.23	1.08	0.00
NIG	0.87	1.00	1.04	0.06	0.83	1.15	1.07	0.00
N	0.76	1.18	1.04	0.05	0.87	1.22	1.09	0.00
t	0.93	1.28	1.06	0.05	0.95	1.14	1.11	0.00
$\hat{\sigma}_2(t) : \varepsilon(t)$	mean	sd	max	min	mean	sd	max	min
HYP	1.31	0.76	1.27	1.42	1.47	0.89	1.39	1.73
NIG	1.25	0.70	1.24	1.29	1.44	0.85	1.43	1.55
N	0.61	0.57	0.61	0.62	0.50	0.00	0.33	0.50
t	0.69	0.78	0.70	0.69	0.51	0.88	0.60	0.50
$\hat{\sigma}_3(t) : \varepsilon(t)$	mean	sd	max	min	mean	sd	max	min
HYP	1.21	1.07	1.16	1.51	1.31	1.20	1.29	1.89
NIG	1.07	1.03	1.12	1.53	1.11	1.15	1.24	2.25
N	1.32	1.19	1.21	1.25	1.58	1.31	1.40	1.40
t	1.22	1.39	1.23	1.50	1.49	1.67	1.38	2.12

Table 3.2: Descriptive statistics of the two criteria for accuracy of estimation: RMAE and RMSE. Two volatility models: local constant (LC) ($\gamma = 0.5$ and $m_0 = 5$) and GARCH(1,1) models are applied to estimate three generated volatility processes. Four kinds of random variables are used to generate the observations: HYP$(2, 0, 1, 0)$, NIG$(2, 0, 1, 0)$, N$(0, 1)$ and $t(6)$.

Next, the sensitivity of the LC and GARCH(1,1) models to jumps in volatility is compared. We introduce a percentage rule to study the sensitivity of the two volatility estimation techniques. The detection speed of the estimated volatility to a sudden jump is measured at a 40%, 50% or 60% level of the jump size. The 40% rule, for example, refers to the number of time steps to reach 40% of the jump size. Table 3.3 provides examples of the detection steps. The GARCH(1,1) process has a naturally fast reaction to jumps in a short interval since it is actually an exponential smoothing process. In general, the LC model needs more time to detect a jump than the GARCH, but the difference is very small. Sometimes the LC model reacts faster than the GARCH(1,1) when adjusting the value of m_0. For example, concerning the jump of σ_{1t} at $t = 300$, the HYPADA needs 4.66 steps on average to detect 50% jump sizes and 6.10 steps to detect 60% jump sizes while the GARCH(1,1) requires 5.17 and 7.27 steps, respectively. In addition, the deviations of these two detections based on the LC method with values of

		$\sigma_1(t=300)$			$\sigma_1(t=600)$		
Model	m_0	40% rule	50% rule	60% rule	40% rule	50% rule	60% rule
GHADA(HYP)	2	3.72(1.7)	4.66(2.9)	6.10(4.7)	5.01(3.2)	6.81(4.7)	8.82(6.5)
GHADA(HYP)	5	4.47(2.2)	5.85(3.3)	7.87(4.8)	5.52(3.0)	7.28(3.9)	9.66(5.6)
HS-ESGH(HYP)		3.63(2.3)	5.17(3.5)	7.27(5.3)	3.42(2.6)	5.28(3.8)	7.69(5.5)
GHADA(NIG)	2	4.35(2.5)	5.64(4.3)	8.88(8.3)	5.81(3.6)	8.09(6.2)	13.22(13.4)
GHADA(NIG)	5	5.92(2.9)	7.94(4.5)	14.69(20.5)	9.25(4.2)	11.98(7.2)	17.89(13.1)
HS-ESGH(NIG)		4.73(3.3)	6.74(4.9)	10.90(8.1)	4.19(2.9)	6.95(5.2)	10.76(9.7)
HS-ADAN	2	3.05(1.6)	3.62(1.9)	4.23(2.4)	3.77(2.1)	5.00(2.7)	6.14(3.5)
HS-ADAN	5	4.29(1.8)	5.17(2.0)	6.19(2.4)	7.24(3.6)	9.49(4.1)	11.19(5.4)
HS-RM		2.56(1.7)	3.27(2.2)	4.27(2.9)	2.27(1.5)	3.11(2.0)	4.00(2.5)
HS-ADAt(6)	2	2.86(1.4)	3.43(1.8)	3.82(2.0)	1.97(1.9)	3.76(2.5)	4.83(2.9)
HS-ADAt(6)	5	3.78(1.7)	4.50(1.9)	5.14(2.2)	1.39(2.1)	6.69(3.8)	8.61(3.9)
HS-ESt(6)		2.42(1.6)	3.03(2.1)	3.76(2.5)	1.64(1.4)	2.47(2.0)	3.25(2.4)
		$\sigma_2(t=400)$			$\sigma_2(t=750)$		
Model	m_0	40% rule	50% rule	60% rule	40% rule	50% rule	60% rule
GHADA(HYP)	2	5.24(3.7)	7.58(4.9)	10.79(7.5)	48.51(33.5)	28.81(19.2)	19.22(12.6)
GHADA(HYP)	5	6.90(3.9)	9.44(5.2)	12.74(9.5)	60.67(45.5)	30.60(21.8)	20.20(13.8)
HS-ESGH(HYP)		4.09(3.0)	7.65(5.0)	12.27(9.2)	73.28(35.4)	30.96(11.4)	17.10(7.7)
GHADA(NIG)	2	6.84(4.3)	10.09(6.9)	15.70(12.9)	29.77(21.7)	18.84(13.6)	9.92(8.4)
GHADA(NIG)	5	8.93(4.6)	12.04(6.7)	18.03(13.5)	39.27(30.1)	22.56(13.8)	13.38(9.0)
HS-ESGH(NIG)		6.63(4.4)	12.71(7.8)	20.49(12.4)	39.49(18.1)	18.17(8.9)	7.94(6.1)

Table 3.3: Mean of the detection steps w.r.t. jumps over 200 simulations. Two methods are implemented to estimate volatility: the local constant (LC) model with $m_0 = 2$ and $m_0 = 5$ and the GARCH(1,1) model. The standard deviations of the detection steps are put in parentheses. Two jumps w.r.t. σ_{1t} at $t = 300$ and $t = 600$ and two jumps w.r.t. σ_{2t} at $t = 400$ and $t = 750$ are considered.

2.9 and 1.7, are smaller than those of the GARCH technique. Meanwhile, we find that the detection speed is slow for a deceased jump. For σ_{2t}, a downward jump from 5% to 1% happens at $t = 750$. The LC model with $m_0 = 2$ needs 19.22 steps on average to detect 60% jump sizes. This number is three times more than that of detection steps for an increased jump with 40% sizes at $t = 400$. This phenomenon results from a low test power in the homogeneity test (3.12), where the squared conditional variance v_I depends on $\theta(t)$ and a larger value of $\theta(t)$ will lead to a low test power.

3.4 Real data analysis

3.4.1 Data set

Two data sets, the DEM/USD exchange rate and a German bank portfolio, are used in the empirical analysis. They are available at FEDC (sfb649.wiwi.hu-berlin.de/fedc).

The daily returns of the exchange rate are calculated from 1979/12/01 to

Data	mean	sd	skewness	kurtosis	ρ_1	ρ_2
exchange rate: $t \in [501, 3719]$						
$r(t)$	-8.30e-05	7.00e-03	-0.07	4.94	0.02	0.01
$\hat{\varepsilon}^{LC}(t)$	-5.24e-03	0.99	-0.01	4.03	0.03	0.02
$\hat{\varepsilon}^{GARCH}(t)$	-7.13e-03	0.99	-0.04	4.38	0.03	0.02
bank portfolio: $t \in [501, 5602]$						
$r(t)$	9.51e-05	1.59e-02	0.28	-8.08	-0.04	-0.03
$\hat{\varepsilon}^{LC}(t)$	-1.13e-02	0.96	0.08	5.18	-0.04	-0.02
$\hat{\varepsilon}^{GARCH}(t)$	1.31e-02	0.99	-0.08	7.38	-0.03	-0.02

Table 3.4: Descriptive statistics for the daily standardized residuals of the exchange rate data and bank portfolio data.

1994/04/01. There are 3719 observations. The bank portfolio data reports the market value of the portfolio held by a German bank (anonymous due to the privacy protection law in Germany). There are 5603 daily observations.

The mean, standard deviation, skewness, kurtosis and the first two autocorrelations ρ_1 and ρ_2 of these two data sets are listed in Table 3.4. All time series are centered around 0 and have leptokurtic distributions as indicated by their kurtoses. Two processes of the standardized returns $\hat{\varepsilon}^{LC}(t) = r(t)/\hat{\sigma}^{LC}(t)$ and $\hat{\varepsilon}^{GARCH}(t) = r(t)/\hat{\sigma}^{GARCH}(t)$ are analyzed according to the two volatility estimation techniques. As discussed before, the standardized returns $\hat{\varepsilon}^{LC}(t)$ and $\hat{\varepsilon}^{GARCH}(t)$, compared to the return series $r(t)$, are expected to be more stationary. Note that these standardized returns still have the heavy-tailed distributional property even after eliminating the influence of the time varying volatility.

The estimated local constant and GARCH(1,1) volatilities of the daily returns are displayed in Figure 3.4. The plots show volatility clustering and graphically reflect the movement pattern of risks. For example, the local constant volatility displays a jump at $t = 3044$ corresponding to a large loss of the exchange rate series. In the German bank portfolio data, the simultaneous movements of loss and volatility are more evident. High risk, i.e. large value of volatility, is observed over the turbulent period as $t \in (3000, 4000)$ and small volatility appears in comparably quiet periods. The GARCH(1,1) technique gives comparable estimations. On the meanwhile, the DEM/USD return series displays a more regular fluctuation due to the liquidation of the exchange rate market. Furthermore, long quiet periods with two extremely turbulent periods are observed in the German bank data, which suggests that large homogeneous intervals will be specified on average

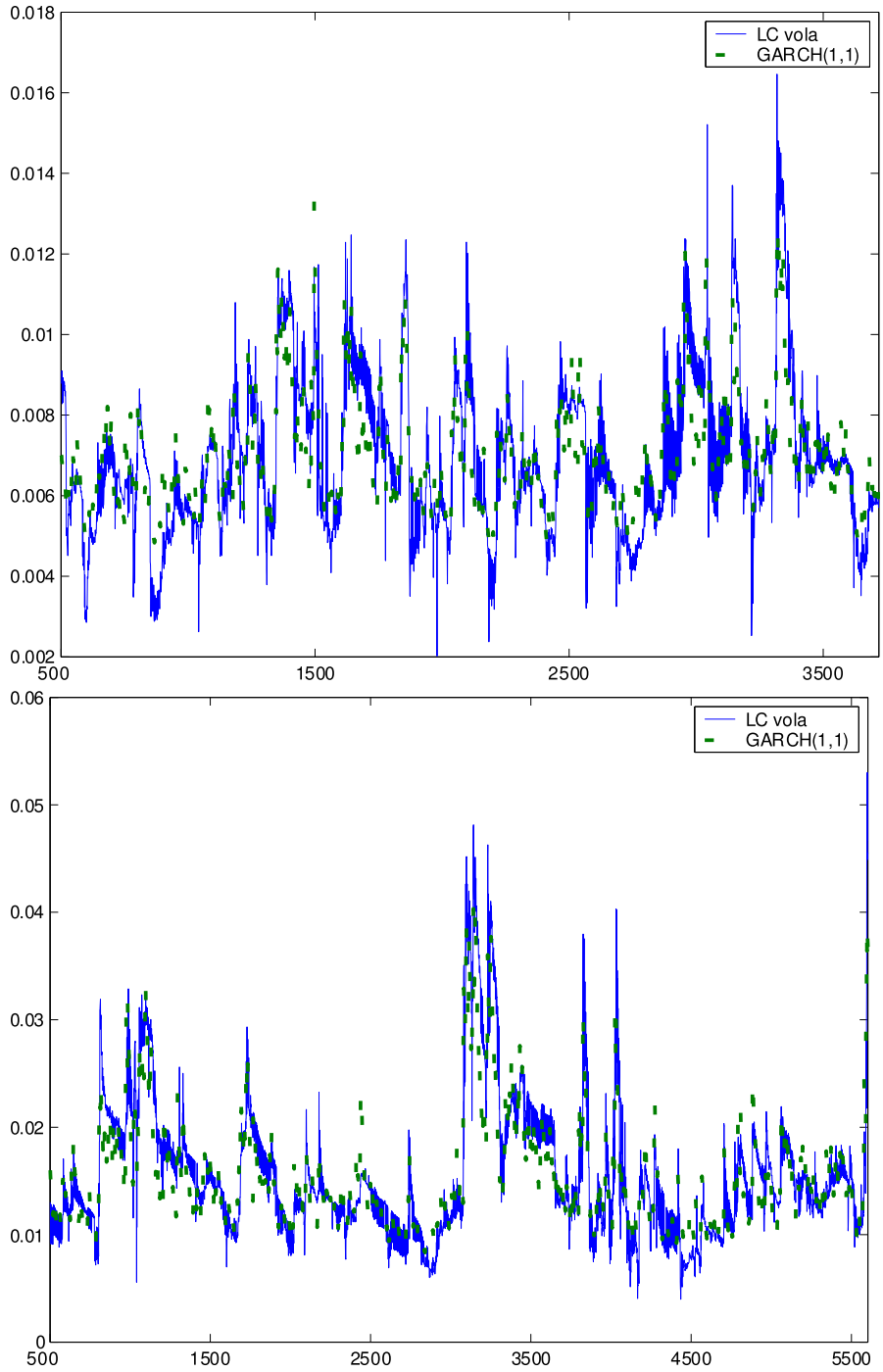

Figure 3.4: Volatility estimation based on the DEM/USD rate (top) and the German bank portfolio (bottom). The parameters in the local constant models are $t_0 = 501$, $m_0 = 5$, $\vartheta' = 1.06$ for the DEM/USD data and $\vartheta' = 1.23$ for the German bank portfolio data.

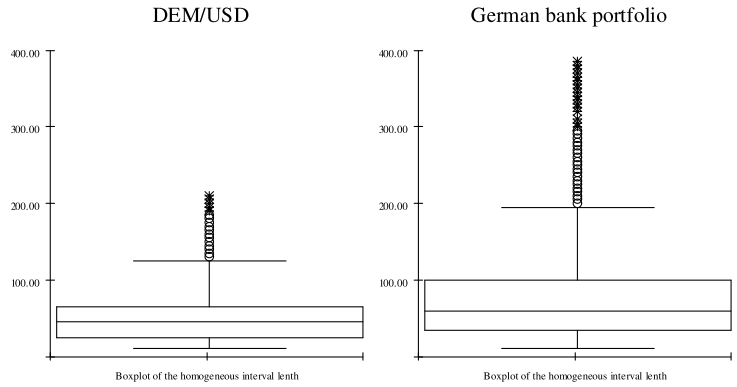

Figure 3.5: Boxplots of the homogeneous interval length w.r.t. the DEM/USD exchange rates (left) and the German bank portfolio data (right).

in the German bank data. Boxplots provide evidence of this suggestion, see Figure 3.5. The means of homogeneous intervals w.r.t. the two data sets are 51.37 (DEM/USD) and 76.42 (German bank), and further, many outliers with large value of interval length (circles or stars with 1.5 or 3 times box length distance from the upper level) are observed in the German bank data.

Given the estimated volatilities, the standardized returns are calculated and used to estimate the distributional parameters of four different assumptions: the HYP, NIG, standard Gaussian and $t(6)$ distributions. The HYP and NIG distributional parameters w.r.t. the local constant and GARCH(1,1) volatilities are listed in Table 3.5. The density estimations under these four

model	exchange rate: $t \in [501, 3719]$				bank portfolio: $t \in [501, 5602]$			
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\mu}$
GHADA(HYP)	1.744	-0.017	0.782	0.012	1.410	0.012	1.0e-13	1.7e-11
GHADA(NIG)	1.340	-0.015	1.337	0.010	0.769	-0.007	0.810	0.019
HS-ESGH(HYP)	1.652	-0.028	0.636	0.021	1.420	0.013	1.5e-15	-8.1e-12
HS-ESGH(NIG)	1.202	-0.026	1.213	0.020	0.879	-0.007	0.880	0.021

Table 3.5: Distributional parameters of the standardized residuals w.r.t. the local constant (LC) volatility and the GARCH(1,1) volatility of the DEM/USD data and the German bank portfolio data.

assumptions are compared graphically. Once again, the nonparametric kernel density estimations implemented in Section 3.1 are used as benchmarks.

Considering the influence of volatility estimation techniques on the standardized returns, the probability densities w.r.t. the local constant (top) and GARCH(1,1) (bottom) techniques are graphed separately in Figure 3.6. Based on the DEM/USD data, the HYP (solid line) and NIG (triangle) models can better describe the distributions of the standardized returns $\varepsilon^{LC}(t)$ and $\varepsilon(t)^{GARCH}$ than the Gaussian (dashed line) and $t(6)$ (dotted line). The Gaussian density underestimates the right tail of the standardized returns whereas the $t(6)$ displays a heavier right tail than the benchmark. This misspecification is enlarged in the log-density comparison (right) in Figure 3.6. Additionally, the standardized returns $\varepsilon^{LC}(t)$ with assumptions of the HYP and NIG distributions match the shape of benchmark better than $\varepsilon^{GARCH}(t)$. Based on the GARCH(1,1) process, all the density estimations deviate from these benchmarks. This weak performance of the GARCH standardization is more obvious in the German bank data, see Figure 3.7. The benchmarks with the GARCH standardization are weakly matched by all the four distributional assumptions. It is possible that the German bank portfolio data is less liquid, for which the local constant model is more suitable, at least theoretically, to capture the movement of the volatility process.

A latent problem of density estimation is whether the distribution of stochastic term is really stationary. A small experiment indicates that the distributional parameters, like volatility, could be time-variant as well. Figure 3.8 shows the HYP-quantile forecasts based on 500 historical standardized returns of the exchange rate for each point in time. It provides evidence that quantiles vary as time passes, especially for extreme probability levels such as $\text{pr} = 0.005$. The same phenomenon holds for the NIG distribution, which is omitted here. If the sample size is small, we could not stick to the assumption that the innovations are identically distributed, although it assumes that the historical observations are i.i.d. as well. Instead, one should update the distributional parameters daily. For example, one may estimate the local distribution based on the previous 500 data points. However as the sample size increases to infinity, the distribution will converge to the unconditional distribution asymptotically. Given the two data sets with large sample size, we assume that all the observations have an identical distribution.

3.4.2 Risk analysis

In this section, we focus on the model selection from the proposals in Table 3.1. Above all, the selected model should be theoretically reasonable and practically tractable to implement. Based on this criterion, we prefer the GHADA model due to its desirable properties discussed before. Another important criterion of model selection is to compare the accuracy of VaR

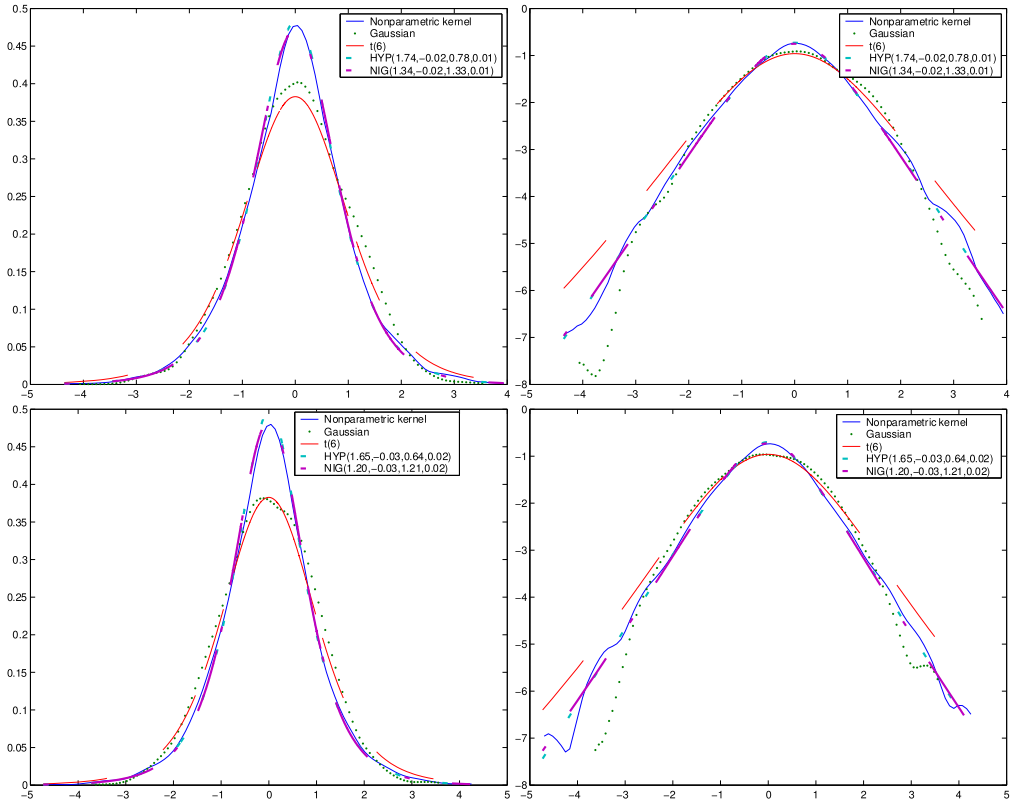

Figure 3.6: Density estimations of the standardized DEM/USD returns under various distributional assumption. The density of the LC-based standardized returns is identified on the top whereas that of the GARCH(1,1)-based standardized returns is illustrated on the bottom. The corresponding log densities are displayed on the right side. The GH distributional parameters are listed in Table 3.5.

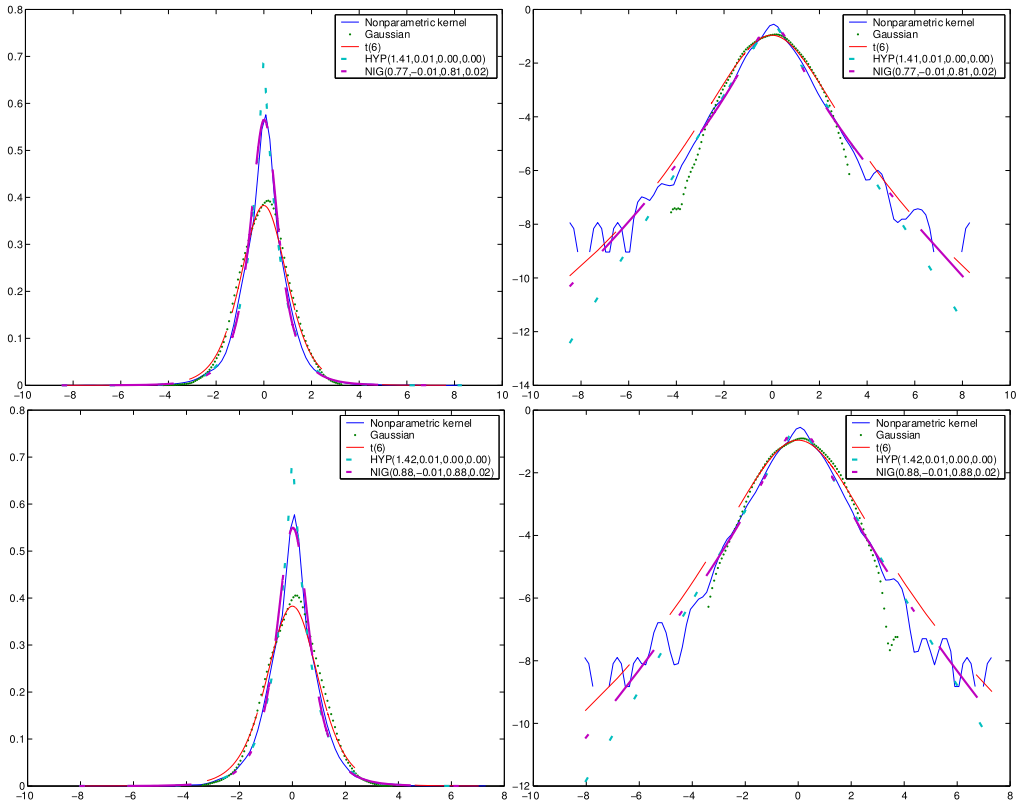

Figure 3.7: Density estimations of the standardized German bank portfolio returns under various distributional assumption. The density of the LC-based standardized returns is identified on the top whereas that of the GARCH(1,1)-based standardized returns is illustrated on the bottom. The corresponding log densities are displayed on the right side. The GH distributional parameters are listed in Table 3.5.

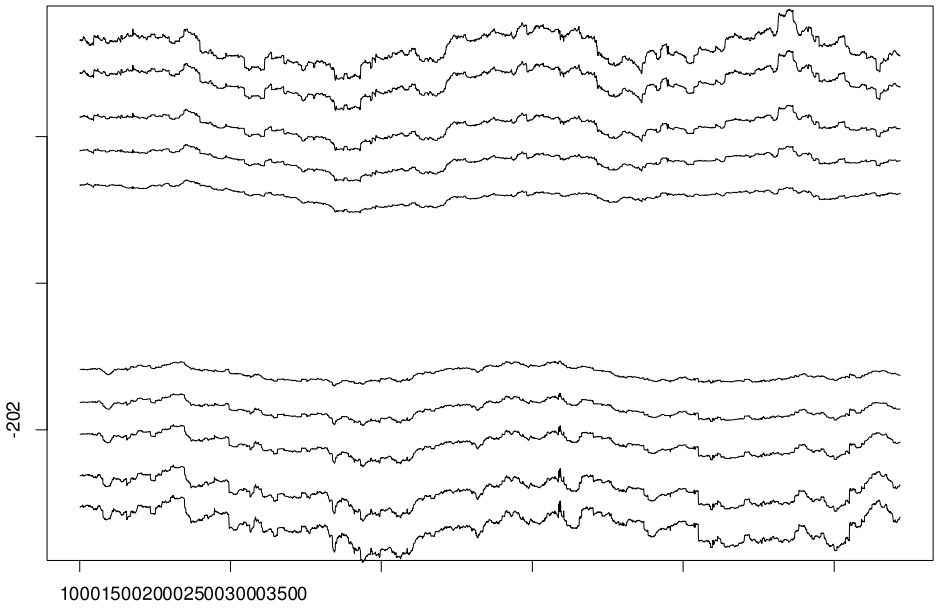

Figure 3.8: Quantiles estimated based on the past 500 standardized returns of the exchange rate. From the top the evolving HYP quantiles for pr = 0.995, pr = 0.99, pr = 0.975, pr = 0.95, pr = 0.90, pr = 0.10, pr = 0.05, pr = 0.025, pr = 0.01, pr = 0.005.

forecasts. The VaR at the probability level pr is forecasted as:

$$\hat{\mathrm{VaR}}_{t,\mathrm{pr}} = -\hat{\sigma}(t)\hat{\mathrm{quantile}}_{\mathrm{pr}}\{\varepsilon(t)\}. \tag{3.18}$$

The quantile is computed according to the distributional estimation. The volatility in the future is unavailable, but empirical studies find that volatility time series appear to have a unit root, see [PG03]. Therefore, the estimated volatility $\hat{\sigma}(t)$ is naturally used as the forecast of tomorrow.

A verified internal model should neither underestimate nor overestimate the market risk. To evaluate the validation of VaR calculation, backtesting concerning exceptions is implemented, see [Chr98]. The exceptions denote returns exceeding the forecasted VaRs. The empirical risk level, the ratio of exceptions in the considered time interval, is compared to the expected risk level $\mathrm{pr} = T^{-1}\sum_{t=1}^{T}\mathbf{1}\{r(t) \leq -\mathrm{VaR}_{t,\mathrm{pr}}\}$, where $\mathbf{1}_t$ denotes the indicator

of exceptions at time point t and T is the total number of forecasts. If the empirical risk level is smaller than pr, it indicates an overestimation of the selected model. In this case, more capital requirements than necessary need to be located in central bank and prospective business opportunities are lost. On the other hand, if the empirical risk level is much larger than pr, i.e. with more exceptions than expected, the multiplicative factor M_f in the market risk charge calculation will increase, see Table 2.1:

$$\text{Risk charge}_t = \max\left(M_f \frac{1}{60}\sum_{i=1}^{60} \text{VaR}_{t-i}, \text{VaR}_t\right) \tag{3.19}$$

Once again, the increase of M_f requires to reserve more capital and reduce the ratio of profits. In our study, the mean of the forecasted VaRs is considered as the amount of risk charge for simplification. The null hypothesis of the backtesting is formulated as:

$$H_0 : \mathsf{E}[N] = T\text{pr} \tag{3.20}$$

with N denotes the number of exceptions. Under H_0 where N is a Binomial random variable with parameters T and pr, the likelihood ratio test statistic can be derived as:

$$\text{LR} = -2\text{log}\{(1-\text{pr})^{T-N}\text{pr}^N\} + 2\text{log}\{(1-N/T)^{T-N}(N/T)^N\}, \tag{3.21}$$

which is asymptotically $\chi^2(1)$ distributed with critical values 3.84 (95%) and 6.63 (99%), [Jor01]. In the recent, another risk measure, the ES, has been used to measure the size of loss:

$$\hat{\text{ES}} = \mathsf{E}\{-r(t)| - r(t) > \hat{\text{VaR}}_{t,\text{pr}}\} \tag{3.22}$$

which is as well calculated here.

Table 3.6 summarizes the forecasted risk charge (RC), the ES and the backtesting results using various models on the base of two data sets: the DEM/USD and the German bank portfolio. As illustrated in Figure 3.6, the HYP and NIG distributional assumptions match the empirical distributional behavior of the standardized DEM/USD returns better than the Gaussian one. The Student-t distribution is although often used in practice, the density comparison shows that this distribution has heavier tails than the empirical one. It is therefore rational to expect that the GH-based models give more accurate VaRs than the other two alternatives. The backtesting exactly shows this case, see Table 3.6. On general, the Gaussian distribution underestimates and the $t(6)$ overestimates the risk. The GHADA method

		DEM/USD data				German bank data			
model	pr	p̂r	LR	RC	ES	p̂r	LR	RC	ES
GHADA(HYP)	0.05	**0.053**	**0.648**	**0.011**	**0.014**	0.048	NAN	0.025	0.034
GHADA(NIG)	0.05	0.054	1.086	0.011	0.014	0.049	NAN	0.025	0.034
HS-ADAN	0.05	0.053	0.417	0.011	0.014	0.045	NAN	0.026	0.035
HS-ADAt(6)	0.05	0.029	*35.379	0.013	0.016	0.032	NAN	0.030	0.039
HS-ESGH(HYP)	0.05	0.053	0.417	0.011	0.015	0.049	NAN	0.025	0.034
HS-ESGH(NIG)	0.05	0.053	0.648	0.011	0.015	**0.050**	NAN	**0.024**	**0.034**
HS-RM	0.05	0.053	0.417	0.011	0.015	0.045	NAN	0.026	0.035
HS-ESt(6)	0.05	0.029	*35.379	0.014	0.017	0.032	NAN	0.030	0.039
GHADA(HYP)	0.01	0.009	0.045	0.018	0.021	0.011	0.173	0.043	0.048
GHADA(NIG)	0.01	**0.010**	**0.001**	**0.018**	**0.021**	0.009	0.516	0.045	0.049
HS-ADAN	0.01	0.015	*6.027	0.016	0.020	0.019	*31.811	0.036	0.044
HS-ADAt(6)	0.01	0.003	*23.609	0.022	0.025	0.006	*10.266	0.049	0.057
HS-ESGH(HYP)	0.01	0.008	0.894	0.019	0.021	0.011	0.306	0.042	0.048
HS-ESGH(NIG)	0.01	0.008	0.894	0.019	0.021	**0.010**	**0.019**	**0.043**	**0.048**
HS-RM	0.01	0.015	*6.027	0.016	0.020	0.019	*31.811	0.036	0.044
HS-ESt(6)	0.01	0.003	*23.609	0.022	0.025	0.006	10.266	0.049	0.057
GHADA(HYP)	0.005	**0.004**	**0.640**	**0.021**	**0.024**	**0.005**	**0.086**	**0.050**	**0.060**
GHADA(NIG)	0.005	**0.004**	**0.640**	**0.021**	**0.024**	0.004	0.509	0.054	0.061
HS-ADAN	0.005	0.010	*13.667	0.018	0.021	0.012	*35.629	0.040	0.047
HS-ADAt(6)	0.005	0.001	*16.164	0.026	0.031	0.003	*4.110	0.058	0.063
HS-ESGH(HYP)	0.005	0.003	3.743	0.022	0.025	0.006	0.459	0.050	0.058
HS-ESGH(NIG)	0.005	0.003	3.743	0.022	0.025	0.005	0.092	0.052	0.062
HS-RM	0.005	0.010	*13.667	0.018	0.021	0.012	*35.629	0.040	0.047
HS-ESt(6)	0.005	0.001	*16.164	0.026	0.031	0.003	*4.110	0.058	0.063

Table 3.6: Backtesting results for the DEM/USD data and the German bank portfolio data. The mean of risk charge and the ES are reported as well. The likelihood ratio (LR) test is asymptotically $\chi^2(1)$ distributed. The critical values are 3.84 (95%) and 6.63 (99%) respectively. * indicates that the risk management model is rejected at 95% confidence level. Notice that NAN for LR is due to the empty set of exceptions.

shows relative good performance. Based on the DEM/USD data, there is no large difference in the accuracy of forecasting between models with the HYP and NIG distributions. On the contrary, the local constant technique performs better than the GARCH(1,1) setup. Given the German bank portfolio data, the similar results are observed. We highlight the "best" model that gives the most accurate empirical probability, the comparable small risk charge on average and ES in the table.

Exemplary time plots of VaRs are displayed in Figure 3.9, by which the GHADA, HS-ADAN and HS-ADAt(6) methods are implemented to forecast VaRs of the DEM/USD returns at $t \in [3000, 3719]$. To be more specific, the volatility is estimated by using the local constant model and the standardized returns are respectively identified in the HYP, Gaussian and Studnet-t distributional frameworks. Two probabilities are considered with pr = 5% and pr = 1%. As expected, the the Gaussian-based VaRs and the GH-based

VaRs locate at similar positions at 5% probability, but the difference becomes evident as the extreme probability is considered. The Student-$t(6)$ distributional assumption although shows heavy tails of the returns, but it overestimates the risks compared to the other two alternatives. Similar phenomena are observed for the German bank portfolio data, which are omitted here.

3.5 Conclusion

In this chapter, we propose a risk management (GHADA) model based on the adaptive moving window average method under GH distributional assumption. Compared to other proposed risk management models in Table 3.1, the GHADA technique gives more accurate VaR forecasts in real data analysis. Some interesting points are summarized:

- Above all, the two subclasses, HYP and NIG, of the GH distribution can better describe the distributional features of the risk factors than the Gaussian and t distributions. Both distributions well match the empirical density, especially the right tail behavior of returns. It leads to precise quantile estimations at extreme risk levels. However we do not have enough evidence to say that one subclass is better than the other, although in the simulation study, the HYP performs better. We consider that the performance of these two subclasses relies on the data set considered. A subjective suggestion is that the NIG is more suitable for less liquid portfolios.

- The local constant technique is an adaptive methods of the popular moving window average estimation. It provides at least as good of a volatility estimation as a GARCH(1,1). The simulation study shows that sometimes the local constant technique provides more accurate volatility estimation on average. Furthermore, the difference between the detection speeds of two techniques is trivial. We want to emphasize here that although the GARCH(1,1) performs well in the simulation study, it is weakened by its theoretical assumption, i.e. the volatility estimation follows a time constant closed form. This influence is illustrated in the German bank portfolio analysis where even the HYP and NIG weakly identify the empirical density of the standardized return based on the GARCH(1,1) technique. On the contrary, the standardized return based on the local constant technique displays a nice fit without problem.

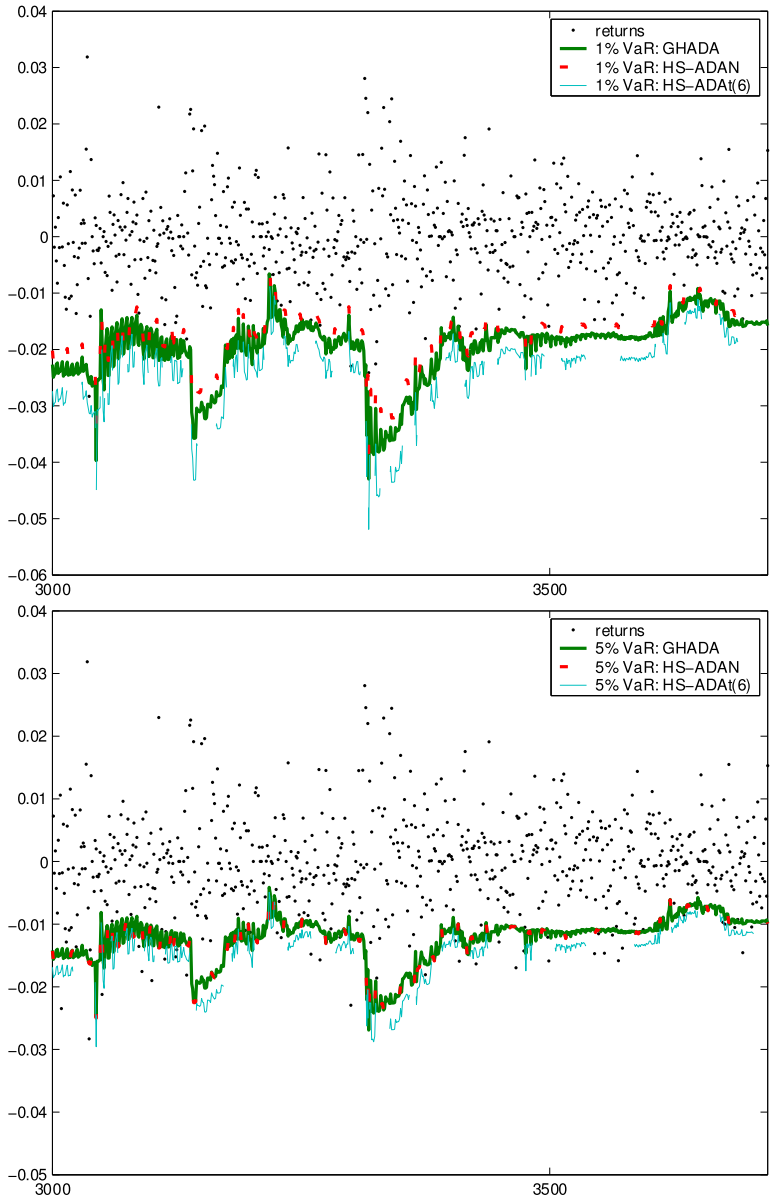

Figure 3.9: Time plots of the VaR forecasts on the base of the DEM/USD returns for the interval $t \in [3000, 3719]$ at pr $= 0.05$ (top) and pr $= 0.01$ (bottom). The volatility is estimated by using the local constant model and the standardized returns are respectively identified in the GH (HYP), Gaussian and Studnet-t distributional frameworks.

3.6 Appendix

For a GH distributed random variable, we have the following lemma:

LEMMA 1 *For every $0 \leq \gamma < 1$ there exists a constant $a_\gamma > 0$ such that*

$$log\, \mathsf{E}[e^{u\zeta_\gamma}] \leq \frac{a_\gamma u^2}{2},$$

where $\zeta_\gamma = (|\varepsilon|^\gamma - C_\gamma)/D_\gamma$ with ε from a GH distribution.

Proof of Lemma 1.
Proof:
Firstly we show that the moment generating function $\mathsf{E}[e^{u\zeta_\gamma}]$ exists for all $u \in \mathbb{R}$.

Suppose that $\mathcal{L}(x) = GH(\lambda, \alpha, \beta, \delta, \mu)$ with the density function f for the transformed variable $y \stackrel{\text{def}}{=} |x|^\gamma$, we have

$$P(y \leq z) = P(-z^{\frac{1}{\gamma}} \leq x \leq z^{\frac{1}{\gamma}}) = \int_{-\infty}^{z^{\frac{1}{\gamma}}} f(x)dx - \int_{-\infty}^{-z^{\frac{1}{\gamma}}} f(x)dx, \quad z > 0$$

Then the density of $y \in (0, \infty)$ is:

$$\begin{aligned} g(z) = \frac{d}{dz}P(y \leq z) &= \gamma^{-1}\{f(z^{\frac{1}{\gamma}})z^{\frac{1}{\gamma}-1} + f(-z^{\frac{1}{\gamma}})z^{\frac{1}{\gamma}-1}\} \\ &= \gamma^{-1}z^{\frac{1}{\gamma}-1}\{f(z^{\frac{1}{\gamma}}) + f(-z^{\frac{1}{\gamma}})\}, \quad z > 0. \end{aligned}$$

Since $f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1}e^{-(\alpha-\beta)x}$ as $x \to \pm\infty$, it follows

$$\begin{aligned} g(z) &\sim \frac{z^{\frac{1}{\gamma}-1}}{\gamma}\{z^{\frac{\lambda-1}{\gamma}}e^{(\beta-\alpha)z^{\frac{1}{\gamma}}} + z^{\frac{\lambda-1}{\gamma}}e^{-(\beta+\alpha)z^{\frac{1}{\gamma}}}\} \\ &= \frac{z^{\frac{\lambda}{\gamma}-1}}{\gamma}\{e^{(\beta-\alpha)z^{\frac{1}{\gamma}}} + e^{-(\beta-\alpha)z^{\frac{1}{\gamma}}}\}, \quad z \longrightarrow \infty \end{aligned}$$

For $\gamma < 1$, it holds that $\int_0^\infty e^{uz}g(z)dz < \infty\ \forall u \in \mathbb{R}$, since

$$\begin{aligned} \lim_{z\to\infty}(\beta-\alpha)z^{\frac{1}{\gamma}} + uz &\to -\infty \quad \forall u \in \mathbb{R} \\ \lim_{z\to\infty} -(\beta+\alpha)z^{\frac{1}{\gamma}} + uz &\to -\infty \quad \forall u \in \mathbb{R} \end{aligned}$$

Since the integration depends only on the exponential part, it holds also that

$$\int_0^\infty z^n e^{uz}g(z)dz = \int_0^\infty \frac{\partial^n}{\partial u^n}(e^{uz})g(z)dz = \frac{\partial^n}{\partial u^n}\mathsf{E}[e^{uy}] < \infty,$$

then it can be shown that the moment generating function and $\log(\mathsf{E}[e^{uy}])$ are smooth. It holds for every $t > 0$,

$$\begin{aligned}\mathsf{E}[e^{uy}] = \mathsf{E}[e^{u|x|^\gamma}] &= \mathsf{E}[e^{u|x|^\gamma}\mathbf{1}(|x| \le t)] + \mathsf{E}[e^{u|x|^\gamma}\mathbf{1}(|x| > t)] \\ &\le e^{ut^\gamma} + \mathsf{E}[e^{|x|ut^{\gamma-1}} I(|x| > t)], \end{aligned} \tag{3.23}$$

Without loss of generality, we assume $\mu = 0$. Further

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1} e^{-(\alpha-\beta)x} \text{ as } x \to \infty,$$

and $\int_y^\infty x^{\lambda-1}e^{-x}dx \sim y^{\lambda-1}e^{-y}$ as $y \to \infty$, [PTVF92].

For an arbitrary but fixed $u \in \mathbb{R}_+$ and $t_0 > 1$ so that $ut^{\gamma-1} < \alpha - \beta$, it holds for all $t \ge t_0$

$$\begin{aligned} f(t) &\le C_1 t^{\lambda-1} e^{(\beta-\alpha)t} \\ \int_{(\alpha-\beta-ut^{\gamma-1})t}^\infty x^{\lambda-1}e^{-x}dx &\le C_2[(\alpha-\beta-ut^{\gamma-1})t]^{\lambda-1} e^{-(\alpha-\beta-ut^{\gamma-1})t} \end{aligned}$$

where $C_1, C_2 > 1$.

Consequently for $t \ge t_0$,

$$\begin{aligned}\mathsf{E}[e^{u|t|^{\gamma-1}x}\mathbf{1}(|x| > t)] &= \int_t^\infty e^{ut^{\gamma-1}x} f(x)dx \le C_1 \int_t^\infty e^{ut^{\gamma-1}x} x^{\lambda-1} e^{-(\alpha-\beta)x} dx \\ &= C_1 \int_t^\infty x^{\lambda-1} e^{-(\alpha-\beta-ut^{\gamma-1})x} dx \\ &= C_1(\alpha-\beta-ut^{\gamma-1})^{-\lambda} \int_{(\alpha-\beta-ut^{\gamma-1})t}^\infty x^{\lambda-1}e^{-x}dx \\ &\le C_1 C_2 t^{\lambda-1} e^{-(\alpha-\beta-ut^{\gamma-1})t} (\alpha-\beta-ut^{\gamma-1}t)^{-1} \end{aligned} \tag{3.24}$$

If u is so large that $t \stackrel{\text{def}}{=} (\frac{\alpha-\beta}{2})^{\frac{1}{\gamma-1}} u^c \ge t_0$ with $\frac{1}{1-\gamma} \le c$, then (3.24) holds true since $ut^{\gamma-1} = (\frac{\alpha-\beta}{2})uu^{c(\gamma-1)} \le \frac{\alpha-\beta}{2} < \alpha - \beta$.

Given $t = (\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}$, we get

$$\mathsf{E}[e^{ut^{\gamma-1}x}\mathbf{1}(|x| > t)] \le \frac{2C_1C_2}{\alpha-\beta}(\frac{\alpha-\beta}{2}u)^{\frac{\lambda-1}{1-\gamma}} e^{-\frac{\alpha-\beta}{2}(\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}}.$$

From which we get

$$\log(\mathsf{E}[e^{ut^{\gamma-1}}\mathbf{1}(x > t)]) \le C_3 + \frac{\lambda-1}{1-\gamma}\log(u) - (\frac{\alpha-\beta}{2})^{\frac{2-\gamma}{1-\gamma}} u^{\frac{1}{1-\gamma}}$$

Further $\log(\mathsf{E}[e^{ut^{\gamma-1}}\mathbf{1}(x > t)])u^{-\frac{1}{1-\gamma}}$ is also bounded for $u \to \infty$. Analogously we can show the bounding of $\log(\mathsf{E}[e^{ut^{\gamma-1}}\mathbf{1}(x < -t)])u^{-\frac{1}{1-\gamma}}$. Therefore for $\gamma < 1$ the whole term $\mathsf{E}[e^{u|x|^\gamma}\mathbf{1}(|x| > t)]u^{-\frac{1}{1-\gamma}}$ is bounded as $u \to \infty$.

Given $t = (\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}$, we have

$$\begin{aligned} e^{ut^\gamma} &= e^{(\frac{\alpha-\beta}{2})^{\frac{\gamma}{1-\gamma}}u^{\frac{1}{1-\gamma}}} \\ u^{-\frac{1}{1-\gamma}}\log(e^{ut^\gamma}) &= (\frac{\alpha-\beta}{2})^{\frac{\gamma}{1-\gamma}} = \text{constant} \end{aligned}$$

Thus $u^{-\frac{1}{1-\gamma}}\log(\mathsf{E}[e^{u|x|^\gamma}]) \leq u^{-\frac{1}{1-\gamma}}[\log(e^{ut^\gamma}) + \log\{E[e^{ut^{\gamma-1}|x|}\mathbf{1}(|x| > t)]\}]$ is bounded for $u \to \infty$, i.e. for a sufficient large u_0 there exist a constant $C_u > 0$ such that

$$\mathsf{E}[e^{u|x|^\gamma}] \leq C_u u^{\frac{1}{1-\gamma}}, \quad u \geq u_0.$$

□

Proof of the Martingale property of $\theta(t)$ Consider a predictable process p_t (such as the volatility $\sigma(t)$ or the local parameter $\theta(t)$) w.r.t. the information set $\mathcal{F}_{t-1}$:

$$\Upsilon_t = \exp\left(\sum_{s=1}^{t} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t} p_s^2\right)$$

$\Upsilon_t$is a supermartingale, since

$$\begin{aligned} \mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) - \Upsilon_{t-1} &= \mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) - \mathsf{E}(\Upsilon_{t-1}|\mathcal{F}_{t-1}) \\ &= \mathsf{E}[\exp\left(\sum_{s=1}^{t} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t} p_s^2\right) \\ &\quad - \exp\left(\sum_{s=1}^{t-1} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t-1} p_s^2\right)|\mathcal{F}_{t-1}] \\ &= \mathsf{E}[\exp\left(\sum_{s=1}^{t-1} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t-1} p_s^2\right) \\ &\quad \{\exp(p_t\zeta_t - a_\gamma/2p_t^2) - 1\}|\mathcal{F}_{t-1}] \\ &= \underbrace{\frac{\exp(p_1\zeta_1)}{\exp(a_\gamma/2p_1)}}_{\leq 1, Lemma1} \cdots \underbrace{\frac{\exp(p_{t-1}\zeta_{t-1})}{\exp(a_\gamma/2p_{t-1})}}_{\leq 1} \cdot \mathsf{E}[\underbrace{\frac{\exp(p_t\zeta_t)}{\exp(a_\gamma/2p_t)}}_{\leq 1} - 1|\mathcal{F}_{t-1}] \\ &\leq 0 \end{aligned}$$

i.e. $\mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) \leq \Upsilon_{t-1}$. By this lemma, we obtain a generalized version of Theorem 3.1 in [MS04a] to the case when ε are from a GH distribution. The statistical properties of $\hat{\theta}_I$ are given in Theorem 1.

Chapter 4

Adaptive risk management 2: LESGH

4.1 Introduction

One of the most difficult tasks in financial engineering and many other fields is to assess, in a meaningful way, volatility of the underlying variable. Given the time-homogeneous volatility model, $R_t = \sqrt{\theta}\varepsilon_t$, for the log-returns R_t, the maximum likelihood estimate (MLE) of the variance θ is the average value of the squared returns R_t^2 provided that the innovations ε_t are standard normal distributed. This model however contradicts a well-known empirical feature of financial time series, i.e. volatility is time varying. Another limitation of such a model is the assumption of normally distributed innovations. The financial data indicates heavy-tailed behavior which is hardly described by the normal distribution. The aim of this paper is to develop some approaches for estimating the volatility parameter θ in the time inhomogeneous situation which are stable w.r.t. non-Gaussian heavy-tailed innovations.

In the recent, numerous articles in the popular journals have detailed methodologies to approximate the MLE w.r.t. a heteroscedastic volatility model, $R_t = \sqrt{\theta_t}\varepsilon_t$. Among many others, the exponential smoothing and its variation are considered as good functional approximations by assigning weights to the squared returns:

$$\theta_t \;=\; \frac{1}{1-\eta}\sum_{m=0}^{\infty}\eta^m R_{t-m-1}^2, \quad \eta\in[0,1).$$

A variation of the exponential smoothing is, for example, the GARCH(1,1)

model introduced by [Eng82] and [Bol86]:

$$\theta_t \;=\; \omega + \alpha R_{t-1}^2 + \beta\theta_{t-1} = \frac{\omega}{1-\beta} + \alpha \sum_{m=0}^{\infty} \beta^m R_{t-m-1}^2$$

Both approximations rely on the smoothing parameter η or β to construct a weighting scheme preceding time point t. Since the most recent observations are considered as including more useful information for future than the others, higher weights are given to them, and lower weights to the very far observations.

To implement the exponential smoothing, one first faces the challenge to choose the smoothing parameter η which can be naturally treated as a memory parameter. The values of η close to one correspond to a slow decay of the coefficients η^m and hence, to a large averaging window, while the small values of η result in a high-pass filtering.

The choice of η is often reasoned by experience. In risk management, for instance, $\eta = 0.94$ has been thought of as an optimized value after it was introduced in the RiskMetrics in 1994. It however arises the question whether the experience based value is really better than others. A study on the exchange rate of the Euro to the Japanese Yen (EURJPY) from 1997/01/02 to 2006/01/05 shows that two estimated volatility processes, with $\eta = 0.752$ and $\eta = 0.94$ respectively, are quite different in value and pattern. Figure 4.1 depicts 100 of these estimated volatilities over the period beginning on 1998/03/05 and ending on 1998/07/24. Obviously, the curve with the recommended smoothing parameter $\eta = 0.94$ is more smooth than the other. But without observing the true volatility, the reason for choosing one versus the other is not clear.

Another more reliable but computationally demanding approach is to choose η by optimizing some objective function such as forecasting errors, see [CFS03].

Notice that the mentioned methods choose one constant smoothing (memory) parameter on the base of available data. It is however reasonable to surmise that there are structure shifts driven by e.g. policy adjustments or economic changes through time, see the example in Figure 4.1. These shifts, in turn, require a change of smoothing parameter to optimize the estimation performance. Table 4.1 presents an example in which the log-returns of the exchange rates EURJPY before and after 1 January 2002 are modelled in the GARCH(1,1) setting. Remember that on this date the Euro banknotes and coins were put into circulation. The MLE of the smoothing parameter β, as expected, is distinct for the two subsets. It changes from 0.75 before the currency transition to 0.95 after the transition. The standard deviation of

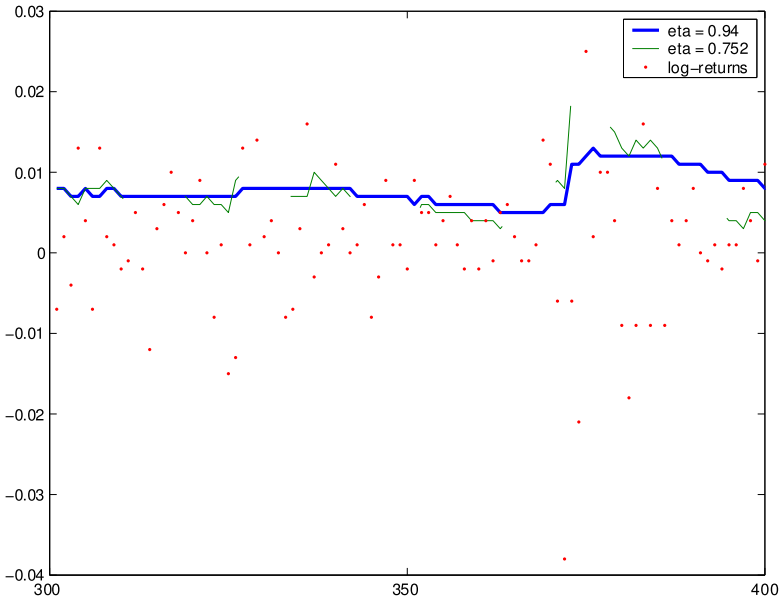

Figure 4.1: Volatility estimates of the EURJPY exchange rates. The exponential smoothing method with constant parameters $\eta = 0.752$ versus $\eta = 0.94$ is applied. The log returns from 1998/03/05 to 1998/07/24 are displayed as dots.

time period	ω	α	β
1997/01/02–2001/12/31	9.07e-06 (2.91e-06)	0.13 (0.03)	0.75 (0.05)
2002/01/02–2006/01/05	5.05e-07 (3.70e-07)	0.02 (0.01)	0.95 (0.02)

Table 4.1: ML estimates of the GARCH(1,1) model. The standard deviation of the estimates are reported in parentheses.

the estimates are put in parentheses. We refer to [BW92] for ML estimation of the parameters.

To alleviate this limitation, we extend the approach to the local version and explain how to choose the smoothing parameter over time. On the meanwhile, a plenty of local model selection methods has been developed, which, to a great extent, avoids serious estimation errors and achieves the best possible accuracy of estimation. These methods automatically investigate the structure shifts and adjust the smoothing parameter to undergo a fast reaction. [MS04b] present an approach to specify local homogeneous interval, by which volatility is approximated by a constant. [FHW95] present a local quasi-likelihood estimation via a local polynomial fitting. We refer to [Spo06] for a detailed description of the local estimation methods.

Among many others, the local model selection (LMS) procedure is a method for selecting a smoothing parameter from a given family. Given a finite set $\eta_1, \ldots, \eta_K$ of the possible values of the memory parameter in the exponential smoothing, one calculates K local MLEs $\{\tilde{\theta}_t^{(k)}\}$ at every time point. The LMS method is to choose one η_k such that its corresponding MLE $\tilde{\theta}_t^{(k)}$ has the best performance in the estimation among the considered set of K estimates. One limitation of the method is that it merely concerns discrete and finite values of the smoothing parameter. The resulting local MLE is in this sense sub-optimal and suffers from a high variability, see [Spo98]. In accordance to the limitation, [BS06] present the spatial aggregation of the local estimate (SSA) by using a convex combination of all the K local estimates instead of choosing one of them. It improves the parameter optimization and produces the local estimate with smaller variability than the LMS method.

The primary aim of the paper is to develop local exponential smoothing, by using the SSA method to optimize the smoothing parameter choice and the accuracy of the local volatility estimate.

The second aim of the paper is to extend the local exponential smoothing in the normal inverse Gaussian (NIG) distributional framework. It has been widely documented that financial series have heavy tails relative to the

Gaussian distribution, see [FHH04]. The heavy tails are typically reduced but not eliminated as the series are standardized by the estimated volatility, see [ABDL01]. The study of [CHJ05] shows that the NIG distribution with four distributional parameters is successful in imitating the distributional behavior of the standardized returns. It is therefore practically interesting to show that the quasi ML estimation is applicable under the NIG distributional assumption.

Moreover, the proposed local volatility estimation methods, combined with the NIG or normal distributional assumption, are implemented to calculate Value-at-Risk (VaR). Since VaR is not a coherent measure of risk exposure, we also examine other risk measures from the viewpoints of regulatory, investors and internal supervisory. On average, the calculations based on the local exponential smoothing and the NIG distributional assumption are favored by investors and supervisory. The VaR values decided by the local volatility estimate and the normal distributional quantile, on the contrary, successfully fulfill the regulatory requirement.

The paper is organized as follows. The local exponential smoothing is described, by which the LMS and SSA methods are used to select the smoothing parameter in Section 4.2. Section 4.3 investigates the choice of parameters involved in the localization. Sensitivity analysis is reported. Later in this section, an alternative parameter tuning is illustrated by minimizing forecasting errors. The quasi ML estimation under the NIG distributional assumption is discussed in Section 4.4. Section 4.5 evaluates the proposed methods based on simulated data. Moreover, risk exposures of two German equities, one US equity and two exchange rates are examined using the proposed local volatility estimation under the normal versus NIG distributional assumption.

4.2 Volatility modeling. Local parametric approach

Let S_t be an observed asset process in discrete time, $t = 1, 2, \ldots$, while R_t defines the corresponding return process: $R_t = \log(S_t/S_{t-1})$. We model this process via the *conditional heteroskedasticity* assumption:

$$R_t = \sqrt{\theta_t}\varepsilon_t\,, \tag{4.1}$$

where ε_t, $t \geq 1$, is a sequence of standardized innovations satisfying

$$\mathsf{E}(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0, \qquad \mathsf{E}(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) = 1$$

with $\mathcal{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$ (σ-field generated by the first $t-1$ observations), and θ_t is the *volatility* process which is assumed to be predictable with respect to $\mathcal{F}_{t-1}$.

In this paper we focus on the problem of filtering the parameter θ_t from the past observations $R_1, \ldots, R_{t-1}$. This problem naturally arises as an important building block for many tasks of financial engineering like Value-at-Risk or Portfolio Optimization.

4.2.1 Local parametric modeling

A *time-homogeneous* (*time-homoskedastic*) model means that θ_t is a constant. For the homogeneous model $\theta_t \equiv \theta$ for t from the given time interval I, and the parameter θ can be estimated using the (quasi) maximum likelihood method. Suppose first that the innovations ε_t are conditionally on $\mathcal{F}_{t-1}$ standard normal. Then the joint distribution of R_t for $t \in I$ is described by the log-likelihood

$$L_I(\theta) = \sum_{t \in I} \ell(y_t, \theta)$$

where $\ell(y, \theta) = -(1/2)\log(2\pi\theta) - y/(2\theta)$ is the log-density of the normal distribution with the parameters $(0, \theta)$ and y_t mean the squared returns, $y_t = R_t^2$. The corresponding maximum likelihood estimate (MLE) maximizes the likelihood:

$$\tilde{\theta} = \operatorname{argmax}_{\theta \in \Theta} L_I(\theta) = \operatorname{argmax}_{\theta \in \Theta} \sum_{t \in I} \ell(y_t, \theta),$$

where Θ is a given parametric subset in $\mathcal{R}_+$.

If the innovations ε_t are not conditionally standard normal, the estimate $\tilde{\theta}$ is still meaningful and it can be considered as a *quasi MLE*.

The assumption of time homogeneity is usually too restrictive if the time interval I is sufficiently large. The standard approach is to apply the parametric modeling in a vicinity of the point of interest t. The localizing scheme is generally given by the collection of weights $W_t = \{w_{st}\}$ which leads to the *local log-likelihood*

$$L(W_t, \theta) = \sum_s \ell(y_s, \theta) w_{st}$$

and to the local MLE $\tilde{\theta}_t$ defined as the maximizer of $L(W_t, \theta)$. In this paper we only consider the localizing schemes with the exponentially decreasing weights $w_{st} = \eta^{t-s}$ for $s \le t$, where η is the given "memory" parameter.

We also cut the weights when they become smaller than some prescribed value $c > 0$, e.g. $c = 0.001$. However, the properties of the local estimate $\tilde{\theta}_t$ are general and apply to any localizing scheme.

We denote by $\tilde{\theta}_t$ the value maximizing the local log-likelihood $L(W_t, \theta)$:

$$\tilde{\theta}_t = \text{argmax}_{\theta \in \Theta} L(W_t, \theta).$$

The volatility model is a particular case of an exponential family, so that a closed form representation for the local maximum likelihood estimate $\tilde{\theta}_t$ and for the corresponding fitted log-likelihood $L(W_t, \tilde{\theta}_t)$ are available, see Polzehl and Spokoiny (2006) for more details.

THEOREM 2 *For every localizing scheme W_t*

$$\tilde{\theta}_t = N_t^{-1} \sum_s y_s w_{st}$$

where N_t denotes the sum of the weights w_{st}:

$$N_t = \sum_s w_{st}.$$

Moreover, for every $\theta > 0$ the fitted likelihood ratio $L(W_t, \tilde{\theta}, \theta) = max_{\theta'} L(W_t, \theta', \theta)$ with $L(W_t, \theta', \theta) = L(W_t, \theta') - L(W_t, \theta)$ satisfies

$$L(W_t, \tilde{\theta}_t, \theta) = N_t \mathcal{K}(\tilde{\theta}_t, \theta) \tag{4.2}$$

where

$$\mathcal{K}(\theta, \theta') = -0.5\{log(\theta/\theta') + 1 - \theta/\theta'\}$$

is the Kullback-Leibler information for the two normal distributions with variances θ and θ': $\mathcal{K}(\theta, \theta') = \mathsf{E}_\theta \, log(I\!P_\theta / dI\!P_{\theta'})$.

Proof:
One can see that

$$L(W_t, \theta) = -\frac{N_t}{2} \log(2\pi\theta) - \frac{1}{2\theta} \sum_s y_s w_{st} \tag{4.3}$$

This representation yields the both assertions of the theorem by simple algebra. □

REMARK 1 *The results of Theorem 2 only rely on the structure of the function $\ell(y, \theta)$ and do not utilize the assumption of conditional normality of the innovations ε_t. Therefore, they apply whatever the distribution of the innovations ε_t is.*

4.2.2 Some properties of the MLE in the homogeneous situation

This section collects some useful properties of the (quasi) MLE $\tilde{\theta}_t$ and of the fitted log-likelihood $L(W_t, \tilde{\theta}_t, \theta^*)$ in the homogeneous situation $\theta_s = \theta^*$ for all s.

We assume the following condition on the set Θ of possible values of the volatility parameter.

($\boldsymbol{\Theta}$) The set Θ is compact and separated away from zero.

First we discuss the case of Gaussian innovations ε_s.

THEOREM 3 (Polzehl and Spokoiny, 2006) *Assume* (Θ). *Let* $\theta_s = \theta^* \in \Theta$ *for* s. *If the innovations* ε_s *are i.i.d. standard normal, then for any* $\mathfrak{z} > 0$

$$I\!P_{\theta^*}(L(W_t, \tilde{\theta}_t, \theta^*) > \mathfrak{z}) \equiv I\!P_{\theta^*}(N_t \mathcal{K}(\tilde{\theta}_t, \theta^*) > \mathfrak{z}) \leq 2e^{-\mathfrak{z}}.$$

Theorem 3 claims that the estimation loss measured by $\mathcal{K}(\tilde{\theta}_t, \theta^*)$ is with high probability bounded by $\mathfrak{z}/N_t$ provided that $\mathfrak{z}$ is sufficiently large. Similarly, one can establish a risk bound for a power loss function.

THEOREM 4 *Assume* (Θ). *Let* y_t *be i.i.d. from* $\mathcal{N}(0, \theta^*)$. *Then for any* $r > 0$

$$\underset{\theta^*}{I\!E} |L(W_t, \tilde{\theta}_t, \theta^*)|^r \equiv \underset{\theta^*}{I\!E} |N_t \mathcal{K}(\tilde{\theta}_t, \theta^*)|^r \leq \tau_r \, .$$

where $\tau_r = 2r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} e^{-\mathfrak{z}} d\mathfrak{z} = 2r\Gamma(r)$. *Moreover, for every* $\lambda < 1$

$$\underset{\theta^*}{I\!E} \exp\{\lambda L(W_t, \tilde{\theta}_t, \theta^*)\} \equiv \underset{\theta^*}{I\!E} \exp\{\lambda N_t \mathcal{K}(\tilde{\theta}_t, \theta^*)\} \leq 2(1-\lambda)^{-1}.$$

Proof:
By Theorem 3

$$\begin{aligned} \underset{\theta^*}{I\!E} |L(\tilde{\theta}_t, \theta^*)|^r &\leq -\int_{\mathfrak{z} \geq 0} \mathfrak{z}^r \, dI\!P_{\theta^*}(L(\tilde{\theta}_t, \theta^*) > \mathfrak{z}) \\ &\leq r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} I\!P_{\theta^*}(L(\tilde{\theta}_t, \theta^*) > \mathfrak{z}) d\mathfrak{z} \leq 2r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} e^{-\mathfrak{z}} d\mathfrak{z} \end{aligned}$$

and the first assertion is fulfilled. The last assertion is proved similarly. □

Theorem 3 can be used for constructing the confidence sets for the parameter θ^*.

THEOREM 5 *Assume* (Θ). *If* $\mathfrak{z}_\alpha$ *satisfies* $2e^{-\mathfrak{z}_\alpha} \leq \alpha$, *then*

$$\mathcal{E}_{t,\alpha} = \{\theta : N_t \mathcal{K}(\tilde{\theta}_t, \theta) \leq \mathfrak{z}_\alpha\}$$

is an α*-confidence set for the parameter* θ^* *in the sense that*

$$I\!P_{\theta^*}(\mathcal{E}_{t,\alpha} \not\ni \theta^*) \leq \alpha.$$

4.2.3 Some properties of the quasi MLE in the homogeneous situation for sub-Gaussian innovations

The assumption of normality for the innovations ε_t is often criticized in the financial literature. Our empirical examples in Section 4.5.2 below also indicate that the tails of estimated innovations are heavier than the normality would imply. The basic result of Theorem 3 and its corollaries can be extended to the case of non-Gaussian innovations under some exponential moment condition. We refer to this situation as the *sub-Gaussian* case. Later these results in combination with the power transformation of the data will be used for studying the heavily tailed innovations, see Section 4.5.2.

THEOREM 6 *Assume* (Θ). *Let the innovations* ε_s *be i.i.d. and fulfill*

$$\log I\!E \exp\{\lambda(\varepsilon_s^2 - 1)\} \leq \kappa(\lambda) \tag{4.4}$$

for all s *and some* $\lambda > 0$. *Then there is a constant* $\mu_0 > 0$ *such that for all* $\theta^*, \theta \in \Theta$

$$\underset{\theta^*}{I\!E} \exp\{\mu_0 L(W_t, \theta, \theta^*)\} \leq 1 \tag{4.5}$$

and

$$I\!P_{\theta^*}(L(W_t, \tilde{\theta}_t, \theta^*) > \mathfrak{z}) \equiv I\!P_{\theta^*}(N_t \mathcal{K}(\tilde{\theta}_t, \theta^*) > \mathfrak{z}) \leq 2e^{-\mu_0 \mathfrak{z}}. \tag{4.6}$$

Proof:
For brevity of notation we omit the subscript t. It holds for $L(W, \theta, \theta^*) = L(W, \theta) - L(W, \theta^*)$

$$2L(W, \theta, \theta^*) = -N\log(\theta/\theta^*) - (1/\theta - 1/\theta^*) \sum_s y_s w_s \,.$$

Under the measure $I\!P_{\theta^*}$, the squared returns y_t can be represented as $y_t = \theta^* \varepsilon_t^2$ leading to the formula

$$\begin{aligned} 2L(W, \theta, \theta^*) &= N\log(\theta^*/\theta) - (\theta^*/\theta - 1) \sum_s \varepsilon_s^2 w_s \\ &= N\log(1+u) - u \sum_s \varepsilon_s^2 w_s \\ &= N\log(1+u) - Nu - u \sum_s (\varepsilon_s^2 - 1) w_s \end{aligned}$$

with $u = \theta^*/\theta - 1$. For any μ such that $\max_s u\mu w_s \le \lambda$ this yields by independence of the ε_s's

$$\begin{aligned}\log \underset{\theta^*}{\mathsf{E}}\{2\mu L(W,\theta,\theta^*)\} &= \mu N\log(1+u) - \mu N u + \sum_s \log \underset{\theta^*}{\mathsf{E}} \exp\{-u\mu w_s(\varepsilon_s^2 - 1)\} \\ &= \mu N\log(1+u) - \mu N u + \sum_s \kappa(-u\mu w_s)\end{aligned}$$

where $\kappa(-u\mu w_s) \le \kappa_0 u^2\mu^2 w_s^2 \le \kappa_0 u^2\mu^2 w_s$. This yields

$$\begin{aligned}\log \underset{\theta^*}{\mathsf{E}}\{2\mu L(W,\theta,\theta^*)\} &\le \mu N\log(1+u) - \mu N u + \sum_s \kappa_0 u^2\mu^2 w_s \\ &= \mu N\{\log(1+u) - u + \kappa_0\mu u^2\}.\end{aligned}$$

The condition (Θ) ensures that $u = u(\theta) = \theta^*/\theta - 1$ is bounded by some constant u^* for all $\theta \in \Theta$. The expression $\log(1+u) - u + \kappa_0\mu u^2$ is negative for all $|u| \le u^*$ and sufficiently small μ yielding (4.6).

Lemma 6.1 from [PS06] implies that

$$\{N_t\mathcal{K}(\tilde\theta_t,\theta^*) > \mathfrak{z}\} \subseteq \{N_t\mathcal{K}(\theta^-,\theta^*) > \mathfrak{z}\} \cup \{N_t\mathcal{K}(\theta^+,\theta^*) > \mathfrak{z}\}$$

for some fixed points θ^+, θ^- depending on $\mathfrak{z}$. This and (4.5) prove (4.6). □

The results of Theorems 4 and 5 extend similarly to the case of sub-Gaussian innovations.

THEOREM 7 *Assume* (Θ) *and (4.4). Then for any* $r > 0$

$$\underset{\theta^*}{\mathsf{E}}|L(W_t,\tilde\theta_t,\theta^*)|^r \equiv \underset{\theta^*}{\mathsf{E}}|N_t\mathcal{K}(\tilde\theta_t,\theta^*)|^r \le \tau_r\,\mu_0^{-r}.$$

Moreover, if $\mathfrak{z}_\alpha$ *satisfies* $2e^{-\mu_0\mathfrak{z}_\alpha} \le \alpha$, *then*

$$\mathcal{E}_{t,\alpha} = \{\theta : N_t\mathcal{K}(\tilde\theta_t,\theta) \le \mathfrak{z}_\alpha\}$$

is an α*-confidence set for the parameter* θ^* *in the sense that*

$$I\!P_{\theta^*}(\mathcal{E}_{t,\alpha} \not\ni \theta^*) \le \alpha.$$

4.2.4 Canonical parametrization

For the proposed stagewise aggregation procedure, it is useful to apply the so called *canonical parametrization*, see e.g. [MN89], for the considered volatility family (P_θ). In this case the *canonical parameter* υ is defined by the

relation $\upsilon = -1/(2\theta)$, so that $\upsilon < 0$. This leads to the following expression for the density of y:

$$p(y, \upsilon) = p(y) \exp\{y\upsilon - d(\upsilon)\}$$

where $d(\upsilon)$ is a convex function defined by $d(\upsilon) = -\frac{1}{2}\log(-2\upsilon)$, $p(y)$ is a nonnegative function.

The local MLE $\tilde{\theta}^{(\eta)}$ yields the corresponding *canonical* estimate

$$\tilde{\upsilon}_t = -1/(2\tilde{\theta}_t).$$

The important results of Theorems 2 and 4 can be restated in terms of the canonical parameter, because they state some properties of the maximum of the log-likelihood function which obviously does not depend on the selected parametrization.

4.2.5 Problem of adaptive estimation

In this paper we focus on the problem of adaptive (data-driven) estimation of the parameter θ_t. We assume that a finite set $\{\eta_k,\, k = 1, \cdots, K\}$ of values of the smoothing parameter is given. Every value η_k leads to the localizing weighting scheme $w_{st}^{(k)} = \eta_k^{t-s}$ for $s \le t$ and to the local ML estimate $\tilde{\theta}_t^{(k)}$:

$$\begin{aligned} \tilde{\theta}_t^{(k)} &= N_k^{-1} \sum_s w_{st}^{(k)} y_s = N_k^{-1} \sum_{m=0}^{M_k} \eta_k^m y_{t-m-1}, \\ N_k &= \sum_s w_{st}^{(k)} = \sum_{m=0}^{M_k} \eta_k^m. \end{aligned}$$

where $M_k = \log c / \log \eta_k - 1$ is the cutting point and guarantees that the weights after M_k are bounded by the prescribed value c, i.e. $\eta_k^{M_k+1} \le c$. It is easy to see that the sum of weights $N_k = \sum_s w_{st}^{(k)}$ does not depend on t, thus we suppress the index t in the notation. The corresponding fitted log-likelihood $L(W_t^{(k)}, \tilde{\theta}_t^{(k)}, \theta)$ reads as

$$L(W_t^{(k)}, \tilde{\theta}_t^{(k)}, \theta) = N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \theta).$$

The local MLEs $\tilde{\theta}_t^{(k)}$ will be referred to as "weak" estimates. Usually the parameter η_k runs over a wide range from values close to one to rather small values, so that at least one of them is "good" in the sense of estimation risk. However, the proper choice of the parameter η generally depends on the variability of the unknown random process θ_s. We aim to construct a data-driven estimate $\hat{\theta}_t$ which performs nearly as good as the best one from this family.

4.2.6 Spatial aggregation of local likelihood estimates (SSA)

The SSA method aggregates all the weak estimates instead of choosing one of them. The procedure is sequential and starts with the estimate $\tilde{\theta}_t^{(1)}$ having the largest variability, that is, we set $\hat{\theta}_t^{(1)} = \tilde{\theta}_t^{(1)}$. At every step $k \geq 2$ the new estimate $\hat{\theta}_t^{(k)}$ is constructed by aggregating the next "weak" estimate $\tilde{\theta}_t^{(k)}$ and the previously constructed estimate $\hat{\theta}_t^{(k-1)}$. Following to [BS06], the aggregation is done in terms of the canonical parameter υ which relates to the natural parameter θ by $\upsilon = -1/(2\theta)$. With $\tilde{\upsilon}_t^{(k)} = -1/(2\tilde{\theta}_t^{(k)})$ and $\hat{\upsilon}_t^{(k-1)} = -1/(2\hat{\theta}_t^{(k-1)})$

$$\begin{aligned} \hat{\upsilon}_t^{(k)} &= \gamma_k \tilde{\upsilon}_t^{(k)} + (1-\gamma_k)\hat{\upsilon}_t^{(k-1)}, \\ \hat{\theta}_t^{(k)} &= -1/(2\hat{\upsilon}_t^{(k)}). \end{aligned}$$

Equivalently one can write

$$\hat{\theta}_t^{(k)} = \Big(\frac{\gamma_k}{\tilde{\theta}_t^{(k)}} + \frac{1-\gamma_k}{\hat{\theta}_t^{(k-1)}} \Big)^{-1}$$

The mixing weights $\{\gamma_k\}$ are computed on the base of the fitted log-likelihood by checking that the previously aggregated estimate $\hat{\theta}_t^{(k-1)}$ is in agreement with the next "weak" estimate $\tilde{\theta}_t^{(k)}$. The difference between these two estimate is measured by the quantity

$$\gamma_k = K_{\mathrm{ag}}(L(W_t^{(k)}, \tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)})/\mathfrak{z}_{k-1}) = K_{\mathrm{ag}}(N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)})/\mathfrak{z}_{k-1})$$

where $\mathfrak{z}_1, \ldots, \mathfrak{z}_{K-1}$ are the parameters of the procedure, see Section 4.3 for more details, and $K_{\mathrm{ag}}(\cdot)$ is the *aggregation kernel*. This kernel monotonously decreases on $\mathcal{R}_+$, is equal to one in a neighborhood of zero and vanishes outside the interval $[0,1]$, so that the mixing coefficient γ_k is one if there is no essential difference between $\tilde{\theta}_t^{(k)}$ and $\hat{\theta}_t^{(k-1)}$ and zero, if the difference is significant. The significance level is measured by the "critical value" $\mathfrak{z}_{k-1}$. In the intermediate case, the mixing coefficient γ_k is between zero and one. The procedure terminates after step k if $\gamma_k = 0$ and we define in this case $\hat{\theta}_t^{(m)} = \hat{\theta}_t^{(k)} = \hat{\theta}_t^{(k-1)}$ for all $m > k$. The formal definition reads as

1. Initialization: $\hat{\theta}^{(1)} = \tilde{\theta}^{(1)}$.
2. Loop: for $k \geq 2$

$$\hat{\theta}_t^{(k)} = \Big(\frac{\gamma_k}{\tilde{\theta}_t^{(k)}} + \frac{1-\gamma_k}{\hat{\theta}_t^{(k-1)}} \Big)^{-1}$$

where the aggregating parameter γ_k is computed as:

$$\gamma_k = K_{\mathrm{ag}}(L(W_t^{(k)}, \tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)})/\mathfrak{z}_{k-1}) = K_{\mathrm{ag}}(N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)})/\mathfrak{z}_{k-1}) \tag{4.7}$$

If $\gamma_k = 0$ then terminate by letting $\hat{\theta}_t^{(k)} = \ldots = \hat{\theta}_t^{(K)} = \hat{\theta}_t^{(k-1)}$.

3. Final estimate: $\hat{\theta}_t = \hat{\theta}_t^{(K)}$.

In a special case of the SSA procedure with the binary γ_k equal to zero or one, every estimate $\hat{\theta}_t^{(k)}$ and hence, the resulting estimate $\hat{\theta}_t$ coincides with one of the "weak" estimates $\tilde{\theta}_t^{(k)}$. This fact can easily be seen by induction arguments. Indeed, if $\gamma_k = 1$, then $\hat{\theta}_t^{(k)} = \tilde{\theta}_t^{(k)}$ and if $\gamma_k = 0$, then $\hat{\theta}_t^{(k)} = \hat{\theta}_t^{(k-1)}$. Therefore, in this situation the SSA method reduces to a kind of *local model selection procedure* (LMS).

The next section discusses in details the problem of the parameter choice and critical values identification for the SSA procedure.

4.3 Parameter choice and implementation details

To run the procedure, one has to specify the setup and fix the parameters of the procedure. The SSA procedure assumes that a growing sequence of values $\eta_1 < \eta_2 < \ldots < \eta_K$ leading to the sequence of "weak" estimates $\tilde{\theta}_t^{(k)}$ is given in advance. The procedure applies to any such sequence for which the following condition is satisfied:

(**MD**) for some constants u_0, u with $0 < u_0 \le u < 1$, the values $N_1, \ldots, N_K$ satisfy

$$u_0 \le N_{k-1}/N_k \le u.$$

Below we present one example of constructing such a set $\{\eta_k\}$ which is used in our simulation study and application examples.

4.3.1 Example of the smoothing parameter set

Every η_k for $k = 1, \cdots, K$ determines the localizing scheme $w_{st}^{(k)} = \eta_k^{t-s}$ for $s \le t$ and $\eta_k^{t-s} > c$ otherwise $w_{st}^{(k)} = 0$. Given a values $\eta_1 < 1$ and $a > 1$, define

$$\frac{N_{k+1}}{N_k} \approx \frac{1-\eta_k}{1-\eta_{k+1}} = a > 1. \tag{4.8}$$

The coefficient a controls the decreasing speed of the variations. The starting value η_1 should be sufficiently small to provide a reasonable degree of localization. Our default values are $a = 1.25$, $\eta_1 = 0.6$, and $c = 0.01$. The total number K of the considered localizing schemes is fixed by the condition that η_K does not exceed the prescribed value $\eta^* < 1$. Section 4.5 presents a simulation study about the influence of the mentioned parameters a, c, α, r on the performance of the procedure.

4.3.2 "Aggregation" kernel

The definition of the mixing coefficients γ_k involves the "aggregation" kernel K_{ag}. Our theoretical study is done under the following assumptions on this kernel:

(**Kag**) The aggregation kernel K_{ag} is monotonously decreasing for $u \in \mathcal{R}_+$, $K_{\mathrm{ag}}(0) = 1$, $K_{\mathrm{ag}}(1) = 0$. Moreover, there exists some $u_0 \in (0,1)$ such that $K_{\mathrm{ag}}(u) = 1$ for $u \le u_0$.

Our default choice is $K_{\mathrm{ag}}(u) = \{1 - (u - 1/6)_+\}_+$ so that $K_{\mathrm{ag}}(u) = 1$ for $u \le 1/6$.

Another choice is the uniform aggregation kernel $K_{\mathrm{ag}}(u) = \mathbf{1}(u \le 1)$. This choice leads the binary mixing coefficients γ_k and hence, to a local model selection procedure.

4.3.3 Critical values

The idea of selecting the critical values is to provide the prescribed performance of the procedure in the simple parametric situation with $\theta_t \equiv \theta^*$.

The way of selecting the critical value is based on the so called "propagation" condition and it can be formulated in a quite general set-up. Recall that the SSA procedure is sequential and delivers after the step k the estimate $\hat{\theta}_t^{(k)}$ which depends on the parameters $\mathfrak{z}_1, \ldots, \mathfrak{z}_{k-1}$. We now consider the performance of this procedure in the simple "parametric" situation of constant volatility $\theta_t \equiv \theta$. In this case the "ideal" or optimal choice between the first k estimates $\tilde{\theta}_t^{(1)}, \ldots, \tilde{\theta}_t^{(k)}$ is the one with the smallest variability, that is, the latest estimate $\tilde{\theta}_t^{(k)}$ whose variability is measured by the quantity N_k, see Theorem 4. Our approach is similar to the one which is widely used in the hypothesis testing problem: to select the parameters (critical values) by providing the prescribed error under the "null", that is, in the parametric situation. The only difference is that in the estimation problem the risk is measured by another loss function. This consideration leads to the following

condition: for all $\theta^* \in \Theta$ and $k = 2, \ldots, K$.

$$\mathop{\mathsf{E}}_{\theta^*} |L(W_t^{(k)}, \tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k)})|^r \equiv \mathop{\mathsf{E}}_{\theta^*} |N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k)})|^r \leq \frac{(k-1)\alpha\tau_r}{K-1}. \tag{4.9}$$

Here τ_r is from Theorem 4, and r and α are the fixed global parameters. The meaning of this condition is that the statistical difference between the adaptive estimate $\hat{\theta}_t^{(k)}$ and the "oracle" estimate $\tilde{\theta}_t^{(k)}$ after the first k steps measured by the left hand-side of (4.9) is bounded by a prescribed constant. As a particular case for $k = K$, the condition (4.9) implies

$$\mathop{\mathsf{E}}_{\theta^*} |N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k)})|^r \leq \alpha\tau_r \,.$$

This means that the final adaptive estimate $\hat{\theta}_t$ is sufficiently close to its non-adaptive counterpart $\tilde{\theta}_t^{(K)}$.

The parameter r in (4.9) specifies the selected loss function. To provide a stable performance of the method and to minimize the Monte Carlo error we suggest the choice $r = 1/2$. The parameter α is similar to the test level parameter, and, exactly as in the testing set-up, its choice depends upon the subjective requirements on the procedure. Small values of α mean that we put more attention to the performance of the methods in the time homogeneous (parametric) situation and such a choice leads to a rather conservative procedure with relatively large critical values. Increasing α would result in a decrease of the critical values and an increase of the sensitivity of the method to the changes in the underlying parameter θ_t at cost of a loss of stability in the time homogeneous situation. For the most of applications, a reasonable range of values α is between 0.2 and 1. Section 4.5 presents a small simulation study which demonstrates the dependence of the critical values on the parameters r and α.

It is important to note that the "hyperparameters" r and α are *global* and their proper choice depends on the particular application while the estimation procedure is *local* and it constructs the estimate $\hat{\theta}_t$ separately at each point. The parameters r and α can be selected in a data driven way by fixing some objective function, e.g., by minimizing the forecast error, see Section 4.3.6, however, we prefer to keep this choice free for the user.

The relation (4.9) gives us $K-1$ inequalities to fix $K-1$ parameters $\mathfrak{z}_1, \ldots, \mathfrak{z}_{K-1}$. However, these parameters only implicitly enter in (4.9) and it is unclear, how they can be selected in a numerical algorithmic way. The next section describes a sequential procedure for selecting the parameters $\mathfrak{z}_1, \ldots, \mathfrak{z}_{K-1}$ one after another by Monte Carlo simulations.

We present an upper bound for the critical values $\mathfrak{z}_k$ which, in particular, claims the existence of the solution of the set of inequalities (4.9) and is useful for the theoretical study.

THEOREM 8 (Belomestny and Spokoiny (2006, Theorem 5.1)) *Assume* (MD) *and* (K_{ag}). *There are three constants* a_0, a_1 *and* a_2 *depending on* u_0, u *and* u_0 *only such that the choice*

$$\mathfrak{z}_k = a_0 + a_1 log\alpha^{-1} + a_2 r log N_k$$

ensures (4.9) for all $k \leq K$.

4.3.4 Sequential choice of the critical values by Monte Carlo simulations

This section presents one way of selecting the critical values $\mathfrak{z}_k$ using Monte Carlo simulations from the parametric model. The procedure utilizes the following technical result.

LEMMA 2 *Let the squared returns* y_t *follow the parametric model with the constant volatility parameter* θ^*, *that is,* $y_t = \theta^* \varepsilon_t^2$. *Then the distribution of the "test statistics"* $L(W_t^{(k)}, \tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)}) = N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)})$ *under* $I\!P_{\theta^*}$ *is the same for all* $\theta^* > 0$.

Proof:
It suffices to notice that under $I\!P_{\theta^*}$ the squared returns y_s fulfill $y_t = \theta^* \varepsilon_t^2$ and for every k, the estimate $\tilde{\theta}_t^{(k)}$ can be represented as

$$\tilde{\theta}_t^{(k)} = N_k^{-1} \sum_s y_s w_{st}^{(k)} = \theta^* N_k^{-1} \sum_s \varepsilon_s^2 w_{st}^{(k)},$$

so that $\tilde{\theta}_t^{(k)}$ is θ^* times the estimate computed for $\theta^* = 1$. The same applies by simple induction argument to the aggregated estimate $\hat{\theta}_t^{(k-1)}$. Therefore, the Kullback-Leibler divergence $\mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k-1)})$ is a function of the ratio $\tilde{\theta}_t^{(k)} / \hat{\theta}_t^{(k-1)}$, in which θ^* cancels. $\square$
This lemma means that it suffices to check the condition (4.9) for one particular θ^*, e.g. for $\theta^* = 1$.

As mentioned previously, the critical values $\mathfrak{z}_k$ are selected successively, starting from $k = 1$. To highlight the contribution of $\mathfrak{z}_1$ in the final risk of the method, we set all the remaining values $\mathfrak{z}_2, \ldots, \mathfrak{z}_{K-1}$ equal to infinity: $\mathfrak{z}_2 = \ldots = \mathfrak{z}_{K-1} = \infty$. Now, for every particular $\mathfrak{z}_1$, we have fixed the whole set of critical values and can run the procedure leading to the estimates $\hat{\theta}_t^{(k)}(\mathfrak{z}_1)$ for $k = 2, \ldots, K$. The value $\mathfrak{z}_1$ is selected as the minimal one for which

$$\mathop{\mathsf{E}}_{\theta^*} |N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k)}(\mathfrak{z}_1))|^r \leq \frac{\alpha \tau_r}{K-1}, \qquad k = 2, \ldots, K. \tag{4.10}$$

Such a value exists because the choice $\mathfrak{z}_1 = \infty$ leads to $\hat{\theta}_t^{(k)}(\mathfrak{z}_1) = \tilde{\theta}_t^{(k)}$ for all k.

Next, with $\mathfrak{z}_1$ fixed in this way, we select $\mathfrak{z}_2$. The method is similar: set $\mathfrak{z}_3 = \ldots = \mathfrak{z}_{K-1} = \infty$ and play with $\mathfrak{z}_2$. Every particular value of $\mathfrak{z}_2$ determines the whole set of critical values $\mathfrak{z}_1, \mathfrak{z}_2, \infty, \ldots, \infty$. The procedure with such critical values results in the estimates $\hat{\theta}_t^{(k)}(\mathfrak{z}_1, \mathfrak{z}_2)$ for $k = 3, \ldots, K$. We select $\mathfrak{z}_2$ as the minimal value which fulfills

$$\mathop{\mathsf{E}}_{\theta^*} |N_k \mathcal{K}(\tilde{\theta}_t^{(k)}, \hat{\theta}_t^{(k)}(\mathfrak{z}_1, \mathfrak{z}_2))|^r \le \frac{2\alpha\tau_r}{K-1}, \qquad k = 3, \ldots, K. \tag{4.11}$$

Such a value exists because the choice $\mathfrak{z}_2 = \infty$ provides a stronger inequality (4.10). We continue this way for all $k < K$. Suppose $\mathfrak{z}_1, \ldots, \mathfrak{z}_{k-1}$ have been already fixed. We set $\mathfrak{z}_k = \ldots = \mathfrak{z}_{K-1} = \infty$ and play with $\mathfrak{z}_k$. Every particular choice of $\mathfrak{z}_k$ leads to the estimates $\hat{\theta}^{(m)}(\mathfrak{z}_1, \ldots, \mathfrak{z}_k)$ for $m = k+1, \ldots, K$ coming out of the procedure with the parameters $\mathfrak{z}_1, \ldots, \mathfrak{z}_k, \infty, \ldots, \infty$. We select $\mathfrak{z}_k$ as the minimal value which fulfills

$$\mathop{\mathsf{E}}_{\theta^*} |N_l \mathcal{K}(\tilde{\theta}_t^{(l)}, \hat{\theta}_t^{(l)}(\mathfrak{z}_1, \ldots, \mathfrak{z}_k))|^r \le \frac{k\alpha\tau_r}{K-1}, \qquad l = k+1, \ldots, K. \tag{4.12}$$

By simple induction arguments one can see that such a value exists and that the final procedure with the such defined parameters fulfills (4.9).

It is also worth mentioning that the numerical complexity of the proposed procedure is not very high. It suffices to generate once M samples from $I\!P_{\theta^*}$ and compute and store the estimates $\tilde{\theta}_t^{(k,m)}$ for every realization, $m = 1, \ldots, M$ and $k = 1, \ldots, K$. The SSA procedure operates with the estimates $\tilde{\theta}_t^{(k)}$ and there is no need to keep the samples themselves. Now, with the fixed set of parameters $\mathfrak{z}_k$, computing the estimates $\hat{\theta}_t^{(k)}$ requires only the finite number of operations proportional to K. One can roughly bound the total complexity of the Monte Carlo study by CMK^2 for some fixed constant C.

4.3.5 Sensitivity analysis. Numerical study

This section presents some numerical results for the proposed procedures. We first specify our set-up and start with the choice of critical values $\mathfrak{z}_k$. Then we illustrate how the SSA procedure works for the simulated and real data.

The parameters $\{\eta_k\}$ defining the weighting scheme $W_t^{(k)}$ are fixed by setting the values c, a, η_1, see Section 4.3.1. We select $c = 0.01$, $a = 1.25$ and $\eta_1 = 0.6$. We also restrict the largest η_K to be smaller than $\eta^* = 0.985$.

k	η_k	M_k	N_k	$\mathfrak{z}_k$ (SSA)	$\mathfrak{z}_k$ (LMS)
1	0.600	9	2.485	0.763	0.273
2	0.680	11	3.095	0.709	0.254
3	0.744	15	3.872	0.655	0.235
4	0.795	20	4.843	0.601	0.216
5	0.836	25	6.045	0.547	0.197
6	0.869	32	7.555	0.493	0.178
7	0.895	41	9.446	0.439	0.159
8	0.916	52	11.806	0.385	0.140
9	0.933	66	14.759	0.331	0.121
10	0.946	83	18.446	0.277	0.102
11	0.957	104	23.051	0.223	0.083
12	0.966	131	28.816	0.169	0.064
13	0.973	165	36.024	0.115	0.045
14	0.978	207	45.029	0.061	0.026
15	0.982	259	56.280		

Table 4.2: Critical values of the SSA and LMS methods w.r.t. the default choice: $c = 0.01$, $a = 1.25$, $\eta_1 = 0.6$, $r = 0.5$ and $\alpha = 1$.

To understand the impact of using a continuous aggregation kernel, we also consider the LMS procedure which comes out of the algorithm for the uniform aggregation kernel $K_{\mathrm{ag}}(u) = \mathbf{1}(u \le 1)$.

For the above defined family of localizing schemes, the critical values $\mathfrak{z}_k$ of the SSA and LMS procedures are fixed by the method from Section 4.3.4. The coefficients $\{\eta_k\}$, the corresponding local window width M_k and the resulting critical values are reported in Table 4.2.

Next few numerical results illustrate the influence of the parameters r, α, a, and c on the critical values $\mathfrak{z}_k$.

The sequences of the critical values $\mathfrak{z}_k$ for the SSA procedure for different combinations of r, α, a, and c are detailed in Table 4.3. We start with the default choice and then slightly vary one parameter fixing the others to the default.

The numerical results can be summarized as follows:

- r (Default choice: $r = 0.5$): The parameter r is the power of the loss function. The default choice $r = 0.5$ leads to the smallest values $\mathfrak{z}_k$ for almost all k among the considered values $r = 0.3, 0.5, 0.7, 1.0$, see Figure 4.2. This, in turns, results in a more sensitive procedure, see Section 4.3.5.

k	default	$r = 0.3$	$r = 0.7$	$r = 1.0$	$\alpha = 0.5$	$\alpha = 1.5$	$c = 0.001$
1	0.763	0.955	1.207	2.003	1.572	0.550	0.696
2	0.709	0.902	1.115	1.849	1.462	0.510	0.647
3	0.655	0.849	1.023	1.695	1.352	0.470	0.598
4	0.601	0.796	0.931	1.541	1.242	0.430	0.549
5	0.547	0.743	0.839	1.387	1.132	0.390	0.500
6	0.493	0.690	0.747	1.233	1.022	0.350	0.451
7	0.439	0.637	0.655	1.079	0.912	0.310	0.402
8	0.385	0.584	0.563	0.925	0.802	0.270	0.353
9	0.331	0.531	0.471	0.771	0.692	0.230	0.304
10	0.277	0.478	0.379	0.617	0.582	0.190	0.255
11	0.223	0.425	0.287	0.463	0.472	0.150	0.206
12	0.169	0.372	0.195	0.309	0.362	0.110	0.157
13	0.115	0.319	0.103	0.155	0.252	0.070	0.108
14	0.061	0.266	0.011	0.001	0.142	0.030	0.059
τ_r	0.598	0.682	0.558	0.549	0.598	0.598	0.598

Table 4.3: Sensitivity analysis: comparison of the critical values $\mathfrak{z}_k$.

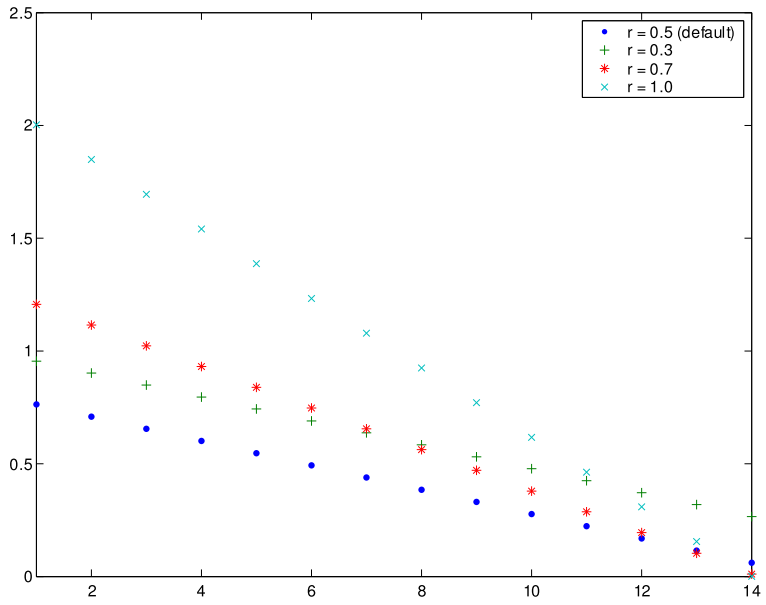

Figure 4.2: Sequences of critical values $\mathfrak{z}_k$ for $r = 0.3$, $r = 0.5$, $r = 0.7$ and $r = 1$.

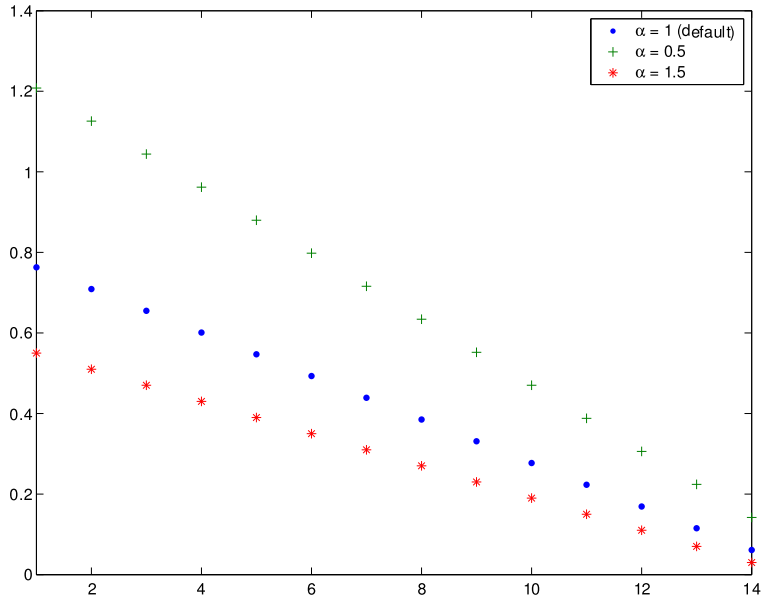

Figure 4.3: Sequences of critical values $\mathfrak{z}_k$ for $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 1.5$.

- α (Default choice: $\alpha = 1$): As already mentioned, the parameter α has the same meaning as the test level. Correspondingly, a decrease of α results in an increase of $\mathfrak{z}_k$ and hence, in a less sensitive procedure. Figure 4.3 illustrates this behaviour.

- a (Default choice: $a = 1.25$): this parameter specifies how dense is the set of possible values η_k. The value of a close to one result in a rather dense set which becomes more and more rare with the increase of a. Therefore, for smaller a-values we have more estimates to select between. This can be helpful for reducing the bias of estimation. However, our theoretical bound for the critical values $\mathfrak{z}_k$ from Theorem 8 indicates that $\mathfrak{z}_k$ increases as K increases. This is confirmed by the numerical results, see Figure 4.4 for a comparison of the critical values $\mathfrak{z}_k$ for the default choice ($K = 15$) and $a = 1.1$ ($K = 34$). Therefore, the use of small a leads to an increase of the critical values and thus, to a less sensitive procedure.

- c (Default choice: $c = 0.01$): The parameter c specifies the cutting point of the exponential smoothing window. As one can expect, this value has only minor influence on the critical values and on the whole procedure. This is in agreement with our numerical results, see

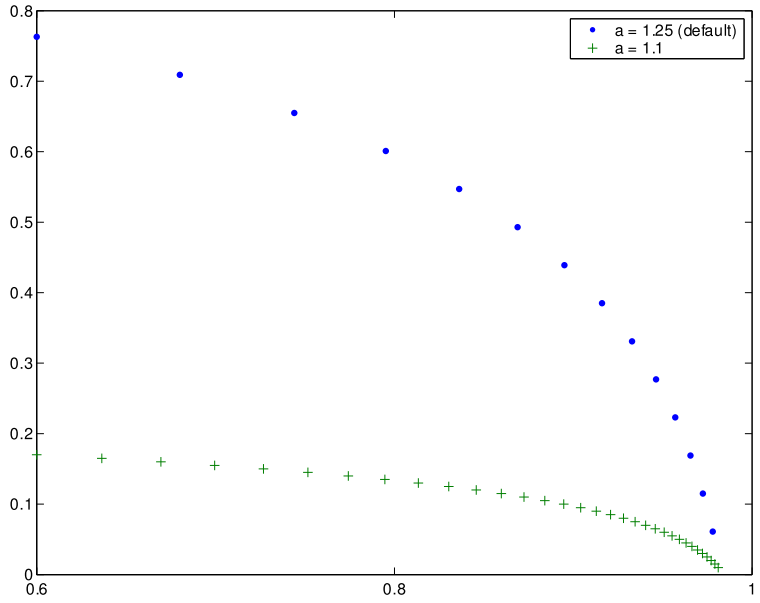

Figure 4.4: Sequences of critical values $\mathfrak{z}_k$ for $a = 1.25$ and $a = 1.1$ w.r.t. the smoothing parameter η_k for $k = 1, \cdots, K$.

Figure 4.5.

4.3.6 Parameter tuning by minimizing forecast errors

The proposed procedure is *local* in the sense that the the adaptation (model selection or aggregation) is performed at every time instant t separately. However, the procedure involves some global parameters like the loss power r or the level α. Their choice can be done in a data-driven way by minimizing the global forecasting error as suggested in [CFS03]. The estimated value $\hat{\theta}_t$ can be viewed as a forecast for the volatility for a short forecasting horizon h. So, a good performance of the method means a relatively small forecast error which is measured as

$$\text{mean } h\text{-step-ahead forecasting errors:} \qquad \sum_{t=t_0}^{T} \frac{1}{h} \sum_{m=0}^{h-1} |y_{t+m} - \hat{\theta}_t|^p$$

for some power $p > 0$.

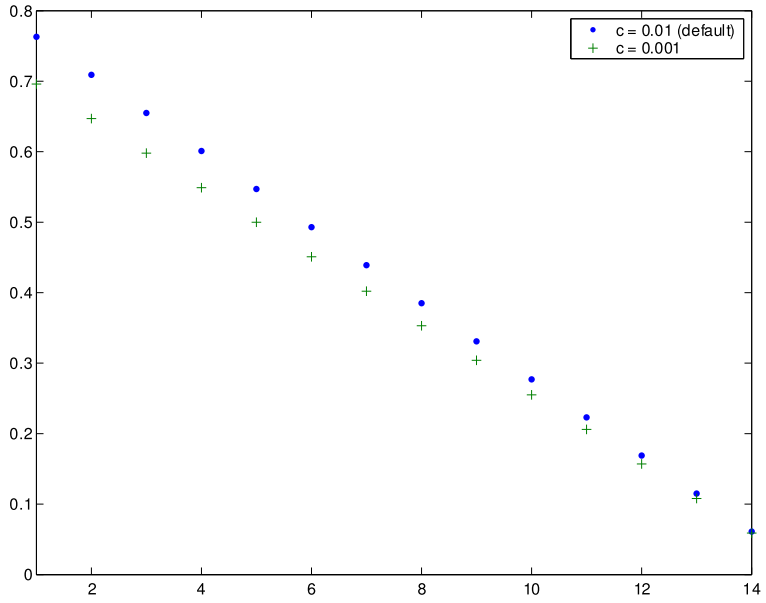

Figure 4.5: Sequences of critical values ϑ_k for $c = 0.01$ and $c = 0.001$.

4.4 Quasi maximum likelihood estimation under normal inverse Gaussian (NIG) distributional assumption

The proposed local exponential smoothing methods and the calculation of the critical values are valid in the Gaussian framework. They can be easily extended to the sub-Gaussian framework considered in Section 4.2.3. However, the financial time series often indicate a heavily tailed behaviour which goes far beyond the sub-Gaussian case. In this section, we extend the methods in the normal inverse Gaussian (NIG) distributional framework which can well describe the heavy-tailed behavior of the real series. Its density is of the form:

$$f_{\mathrm{NIG}}(\varepsilon;\alpha,\beta,\delta,\mu) = \frac{\alpha\delta}{\pi}\frac{K_1\left(\alpha\sqrt{\delta^2+(\varepsilon-\mu)^2}\right)}{\sqrt{\delta^2+(\varepsilon-\mu)^2}}\exp\{\delta\sqrt{\alpha^2-\beta^2}+\beta(\varepsilon-\mu)\},$$

where the distributional parameters fulfill $\mu \in \mathbb{R}$, $\delta > 0$ and $|\beta| \leq \alpha$. By $K_\lambda(\cdot)$ we denote here the modified Bessel function of the third kind which is of the form:

$$K_\lambda(y) = \frac{1}{2}\int_0^\infty y^{\lambda-1}\exp\{-\frac{y}{2}(y+y^{-1})\}\,dy.$$

We refer to [Pra99] for a detailed description of the NIG distribution. One can easily see that the exponential moments $\mathsf{E}\{\exp(\lambda\varepsilon_t^2)\}$ of the squared NIG innovations ε_t^2 do not exist. Hence, the results of Section 4.2.3 do not apply to NIG innovations. From the practical point of view, it is well known that the ES volatility estimate $\tilde{\theta}_t$ is rather unstable under presence of heavy-tailed outliers. This feature makes questionable the use of the ES estimate in the risk management and leads to the problem of constructing the other procedure which are more robust to heavy-tailed data. One popular method for achieving this feature is to apply a power transformation to the underlying process. [BC64] stimulated the application of power transformation to non-Gaussian variables to obtain another distribution more close to the normal and homoscedastic assumption. Here we follow this way and replace the squared returns y_t by their p-power to provide that the resulting "observations" $y_{t,p} = y_t^p$ have exponential moments. One easily gets

$$\mathsf{E}\{y_{t,p} \mid \mathcal{F}_{t-1}\} \;=\; \mathsf{E}\{y_t^p \mid \mathcal{F}_{t-1}\} = \theta_t^p \,\mathsf{E}\,|\varepsilon_t|^{2p} = \theta_t^p C_p = \vartheta_{t,p} \qquad (4.13)$$

where $C_p = \mathsf{E}\,|\varepsilon_t|^{2p}$ is a constant and relies on p and the distribution of the innovations ε_t which is assumed to be NIG. Note that the equation (4.13) can be rewritten as

$$y_{t,p} = \vartheta_{t,p}\varepsilon_{t,p}^2$$

where the "new" standardized squared innovations

$$\varepsilon_{t,p}^2 = y_{t,p}/\vartheta_{t,p} = y_t^p/(C_p\theta_t^p)$$

satisfy $\mathsf{E}\{\varepsilon_{t,p}^2 \mid \mathcal{F}_{t-1}\} = 1$. [CHJ05] proved that $\mathsf{E}\exp(\lambda|\varepsilon_t|^{2p})$ is bounded for NIG case as $0 \le p < 1/2$. This enables us to apply the proposed SSA procedure to the transformed data $y_{t,p}$ to estimate the parameter ϑ_t. An important question for this application is the choice of parameters of the method, especially of the critical values $\mathfrak{z}_k$. The formal application of the approach of Section 4.3.3 requires to use for the Monte Carlo simulations the underlying NIG distribution for the innovations ε_t. This means that one first simulates the NIG data y_t under the hypothesis of time-homogeneous volatility in the form $y_t = \theta^*\varepsilon_t$ with NIG ε_t and then computes the transformed data $y_{t,p}$ which are used for calculating the "weak" estimates $\tilde{\vartheta}_{t,p}^{(k)}$. However, this approach would require the exact knowledge of the parameters of the NIG distribution which is difficult to expect in real life situation. Note that the use of power transformation makes the distribution of the "new" innovations $\varepsilon_{t,p}$ close to the Gaussian case. This suggests to first apply the critical values $\mathfrak{z}_k$ computed for the Gaussian case. Below in this section we

show that the critical values $\mathfrak{z}_k$ are quite stable with respect to the parameters of the NIG distribution and the approach based on the use of Gaussian $\varepsilon_{t,p}$ in the Monte Carlo simulations works well and delivers almost the same results as if the true NIG distribution for the ε_t's would be utilized.

The adaptive procedure delivers the estimate $\hat{\vartheta}_{t,p}$ of the parameter $\vartheta_{t,p}$. To get the estimate of the original parameter θ_t from the relation (4.13) we need to know the constant C_p which generally depends upon the parameters of the NIG distributions. We suggest two ways to fix this constant. One is based on the fact that the standardized innovations $\varepsilon_t^2 = y_t/\theta_t$ should satisfy $\mathsf{E}\,\varepsilon_t^2 = 1$. The estimates $\hat{\theta}_t = \hat{\vartheta}_{t,p}^{1/p}/C_p^{1/p}$ leads to the estimated squared innovations $\tilde{\varepsilon}_t^2 = y_t/\hat{\theta}_t = C_p^{1/p} y_t/\vartheta_{t,p}^{1/p}$, so that C_p can be obtained from the equation

$$n^{-1} C_p^{1/p} \sum_{t=t_0}^{t_1} \frac{y_t}{\vartheta_{t,p}^{1/p}} = 1.$$

where $n = t_1 - t_0 + 1$ means the number of points in which the estimation has been done. The problem with this approach is that the presented sum of $y_t/\vartheta_{t,p}^{1/p}$ is quite sensitive to extreme values of y_t and even one or two outliers can dramatically destroy the resulting estimate.

The other method of fixing the constant C_p is based on the proposal of Section 4.3.6 to minimize the mean forecasting error. Namely, we define the value C_p in a way to minimize

$$\sum_{t=t_0}^{t_1} \frac{1}{h} \sum_{m=0}^{h-1} |y_{t+m} - \hat{\theta}_t|^p = \sum_{t=t_0}^{t_1} \frac{1}{h} \sum_{m=0}^{h-1} |y_{t+m} - \hat{\vartheta}_{t,p}^{1/p}/C_p^{1/p}|^p$$

After the constant C_p is estimated one can use the estimated returns $\tilde{\varepsilon}_t$ for fixing the NIG parameters which will be used for our risk evaluation.

In this study, we choose $p = 0.25$.

4.5 Numerical study

In this section, we first apply the ES, LMS and SSA methods to simulated data with piecewise constant volatility and compare the estimation errors of these estimates. Among them, the SSA estimates are sensitive to structure shifts and fast react to downward volatility jumps. Furthermore, the SSA method on average presents the best accuracy of estimation. In the empirical study, we estimate the local volatility of real financial instruments and use them to calculate risk measures.

4.5.1 Simulation study

This section aims to illustrate the performance of the proposed adaptive procedure relative to the well established non-adaptive ES estimate with the "optimized" parameter $\eta = 0.94$. We consider two versions of the SSA procedure: one with the default parameter set and the other one with the uniform kernel K_{ag} which does a model selection and therefore, referred to as LMS.

In the simulation study, we generate 100 stochastic processes driven by the hidden Markov model: $R_t = \sqrt{\theta_t}\varepsilon_t$ with ε_t either standard normal or NIG with parameters $\alpha = 1.340$,$\beta = -0.015$,$\delta = 1.337$,$\mu = 0.010$. These NIG parameters are in fact the ML estimates of the Deutsche Mark to the US Dollar daily rates from 1979/12/01 to 1994/04/01 available at the FEDC (http://sfb649.wiwi.hu-berlin.de/fedc). The volatility process is a discrete Markov chain with 7 states : $0.2, 0.25, 0.3, 0.4, 0.5, 0.7$ and 1. The transition matrix is not shown here. The generated volatility series is displayed in Figure 4.6. The sample size of the stochastic processes is $T = 1000$. The first 300 observations are reserved as a training set for the very beginning volatility estimation. The volatility of the generated processes is estimated by using the ES, LMS and SSA methods.

Two criteria are used to measure the estimation errors, the sum of the absolute errors (AE) and the ratio of the AEs driven by the local methods to the AE of the ES:

$$\begin{aligned} \text{AE} &= \sum_{t=301}^{N} |\sqrt{\hat{\theta}_t} - \sqrt{\theta_t}| \\ \text{RAE} &= \frac{\text{AE}_{\underline{\text{method}}}}{\text{AE}_{\text{ES}}}. \end{aligned}$$

The estimated variance process of one standard normal distributed realization, i.e. $\varepsilon_t \sim \text{N}(0, 1)$, is displayed in Figure 4.6. The diagram on the top shows that the SSA estimate in general is more close to the true value than the ES one. It successfully follows the movement of the generated variance. Once a downward jump happens for example, the local exponential smoothing reacts much faster than the ES. The difference between the SSA and the LMS is however not significant in the plot. It is illustrated in the comparison of the realized AEs in the second diagram, by which the SSA has a smaller AE than the LMS, and the AEs of the ES method with different smoothing parameters in the sequence of $\{\eta_k\}$ are displayed as well. The best performance of the ES is realized as $\eta = 0.895$ that corresponds to the 7-th smoothing window, see Table 4.2. It shows that even the optimized ES has a larger estimation bias than the local adaptive methods.

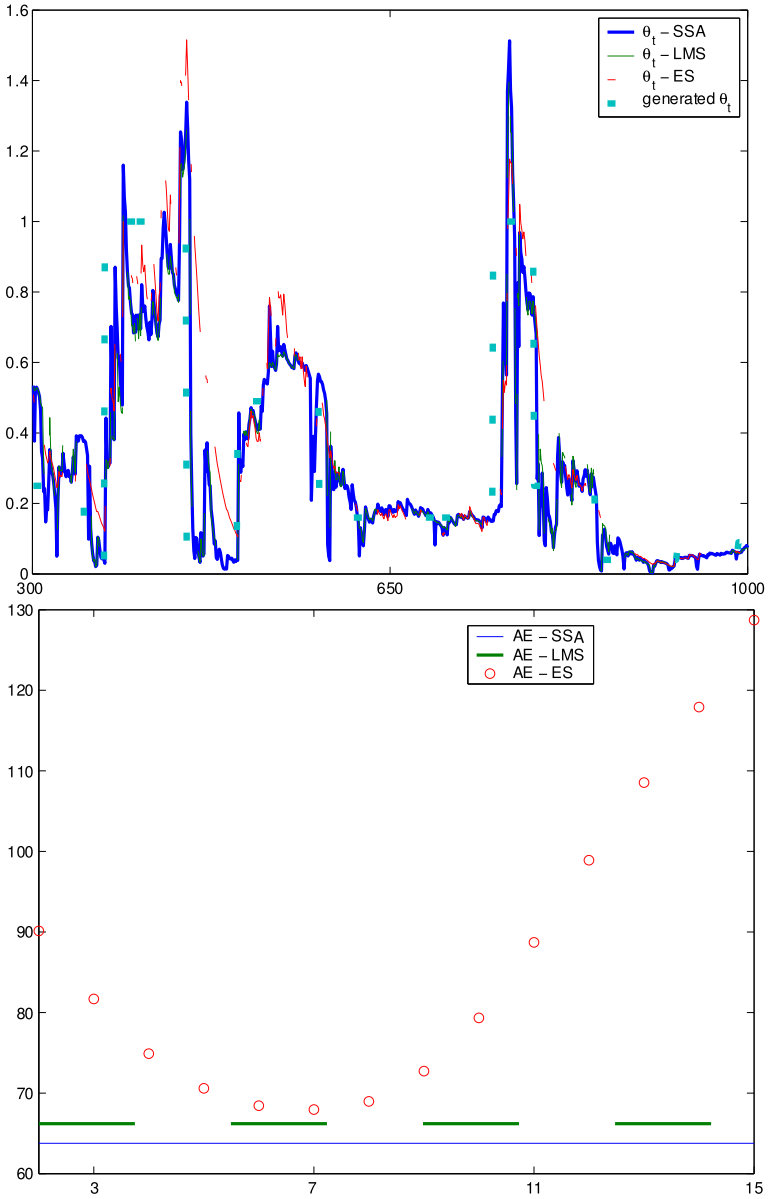

Figure 4.6: Estimation based on one realized simulation data with $\varepsilon_t \sim \mathrm{N}(0,1)$. The ES ($\eta = 0.94$), LMS and SSA estimates and the generated variance process are depicted on the top. The absolute errors of the LMS and SSA estimates are compared with the ES estimates w.r.t. $\{\eta_k\}$ for $k = 2, \cdots, 15$.

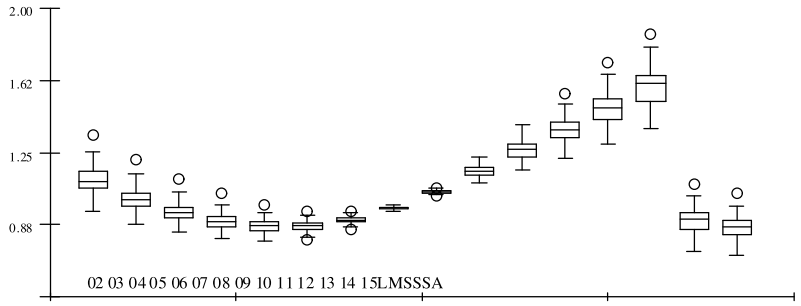

Figure 4.7: The boxplots of the RAEs of the SSA, LMS and ES with η_k for $k = 2, \cdots, K$.

	default choice			r			α		a	c
	SSA	LMS	ES	0.3	0.7	1.0	0.5	1.5	1.1	0.001
AE	69.95	72.94	81.54	73.79	71.92	77.46	76.32	69.63	69.39	69.96
RAE	0.859	0.905	1.000	0.905	0.882	0.950	0.936	0.854	0.851	0.858

Table 4.4: Average estimation errors of the 100 simulation data sets with $\varepsilon_t \sim \mathrm{N}(0, 1)$, by which various estimation methods and the different parameters analyzed in the sensitivity analysis and used in the SSA estimation are considered. SSA ^{2} means the SSA with the critical values based on forecasting errors ($h = 1$). In the ES, $\eta = 0.94$ is applied.

Furthermore, the SSA estimates through the 100 normal distributed simulations have the smallest average AE with the value of 69.946 than the others (LMS: 72.943 and ES 81.538). The corresponding average RAE is 85.924% indicating a 16% improvement versus the ES. The boxplots of the RAEs corresponding to the SSA, LMS and ES with various smoothing parameters, i.e. $\eta = \eta_k$ for $k = 2, \cdots, 15$, are displayed in Figure 4.7. It shows that both the mean and the variation of the RAEs in the SSA method are smaller than those of the LMS method. Figure 4.8 shows the average estimation errors through the simulations. As more simulation results are considered, the average RAE of the SSA is decreasing and it is always smaller than the average RAE of the LMS. A convergence of the RAEs is observed.

Table 4.4 summarizes the estimation errors w.r.t. different methods and the four parameters analyzed in Section 4.3.5. The comparison of the RAEs relative to the default SSA reasons the default choice.

Given the simulated heavy-tailed data, we follow the consequence ex-

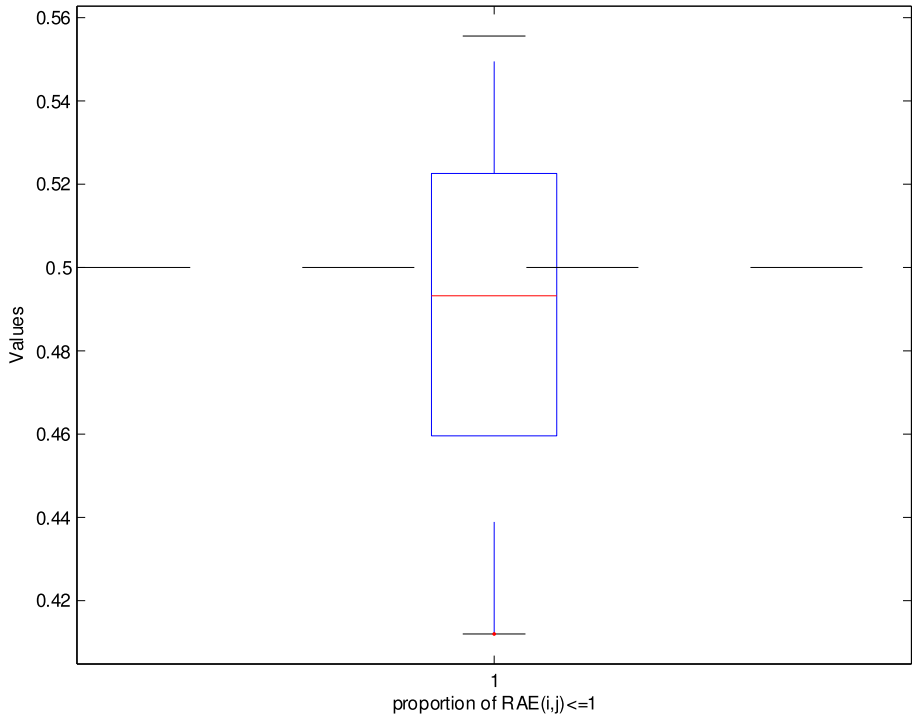

Figure 4.8: The average RAEs of the SSA (blue and solid curve) and LMS (cyan and dotted curve) estimates through the 100 simulations.

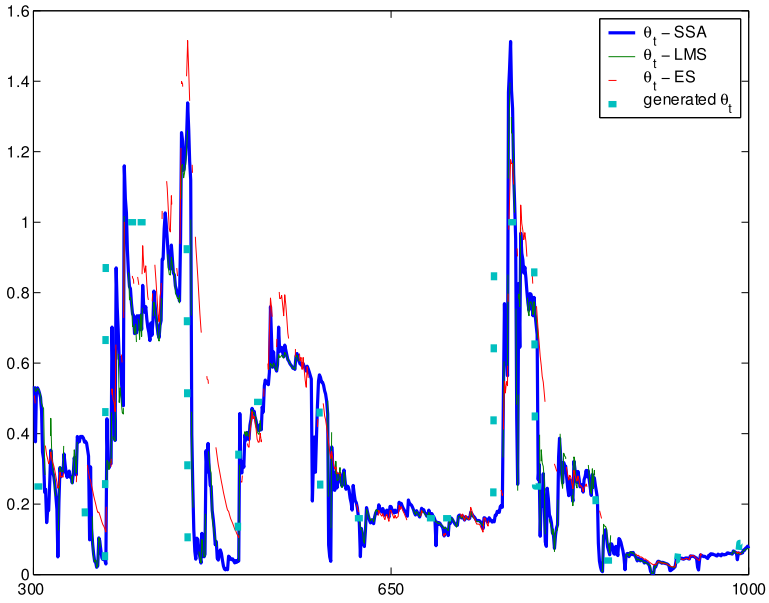

Figure 4.9: Estimation based on one realized simulation data with $\varepsilon_t \sim \text{NIG}(1.340, -0.015, 1.337, 0.010)$. The ES ($\eta = 0.94$) and SSA ($p = 0.25$) estimates and the generated variance process are depicted.

plained in Section 4.4 and first apply the critical values $\mathfrak{z}_k$ computed for the Gaussian case. The results are then used to estimate the distributional parameters of the NIG innovations. Consequently, the critical values w.r.t. the NIG distribution are calculated in the Monte Carlo study and applied to adaptive estimate $\hat{\vartheta}_{t,p}$. Finally the original parameter θ_t are calculated. One realization is displayed in Figure 4.9. As same as the Gaussian case, the SSA estimates are more sensitive to a structure shift than the ES estimates. Given values of p changed from 0.1 to 0.75, the sensibility of the SSA versus the ES is decreasing based on the 100 simulations. The resulting RAEs of the two estimation methods are reported in Table 4.5. At the default case with $p = 0.25$, the RAE is 0.56, indicating that the SSA is almost double accurate than the ES. As the value of p increases to 0.75, the SSA has even a worse accuracy than the ES. As expected, the SSA estimation based on the Gaussian case works well and delivers almost the same accuracy as the NIG case. As $p = 0.25$ for example, the average RAE of the estimates using the Gaussian critical values is 0.58, which is very close to the value based on the NIG critical values. We also list the average value of the true C_p and its estimates over the simulations. These values are very close.

p	RAE	AE(SSA)	AE(ES)	C_p	$\hat{C}_p$
0.25^*	0.58	116.00	201.35	0.78	0.76
0.25	0.56	112.91	201.35	0.78	0.72
0.10	0.48	131.48	276.28	0.90	0.84
0.50	0.77	100.82	130.85	0.60	0.64
0.75	1.08	102.64	95.09	0.47	0.63

Table 4.5: Average estimation errors of the 100 NIG data sets w.r.t. different values of p, by which $p = 0.25$ is default choice in our study. The row $p = 0.25^*$ is based on the critical values computed for the Gaussian case. The true $C_p = \mathsf{E}\,|\varepsilon_t|^{2p}$ and the estimated $\hat{C}_p$ are reported. In the ES, $\eta = 0.94$ is applied.

4.5.2 Application to risk analysis

We consider log-returns of three equity assets Microsoft (MC), Volkswagen (VW), Deutsche Bank (DB) with daily closed price from 2002/01/01 to 2006/01/05 (972 observations) and two exchange rates: EUR/USD (EURUSD) and EUR/JPY (EURJPY) from 1997/01/02 to 2006/01/05 (2332 observations). The data sets have been provided by the financial and economic data center (FEDC) of the Collaborative Research Center 649 on Economic Risk of the Humboldt-Universität zu Berlin.

Among them, the EURUSD series is close to the normal distribution with a kurtosis 3.903, but the series is asymmetric with the value of skewness -0.277. The other series exhibit heavy tails with a large value of kurtosis relative to 3. We model the log-returns of these underlying series by assuming that the innovations are either NIG or Gaussian distributed:

$$R_t = \sqrt{\theta_t}\varepsilon_t, \quad \text{or} \quad \varepsilon_t \sim \mathrm{N}(0,1) \quad \text{or} \quad \varepsilon_t \sim \mathrm{NIG} \tag{4.14}$$

Without loss of generality, the drift of R_t is set to 0. The time-varying volatility is estimated using the proposed ES, LMS and SSA methods. The NIG distributional parameters of the standardized returns are identified using the ML estimation based on the estimated returns $\tilde{\varepsilon}_t = R_t/\sqrt{\hat{\theta}_t}$. These parameters are assumed to be time-homogeneous. In other words, the volatility explains all the possible structural shifts. The KPSS test of stationarity w.r.t. the standardized returns is not rejected at the 90% confidence level, see Table 4.6.

Risk exposures of each financial series are measured from the viewpoints of regulator, investors and internal supervisor.

Data	method	mean	s.d.	skewness	kurtosis	KPSS
MC return		0.000	0.012	0.655	10.965	0.025
$\hat{\varepsilon}_t$	SSA	0.001	1.235	0.261	10.494	0.059
	LMS	-0.004	1.204	0.065	10.173	0.085
	ES	-0.003	1.071	0.545	12.492	0.036
VW return		-0.001	0.019	0.242	6.779	0.092
$\hat{\varepsilon}_t$	SSA	-0.063	1.150	0.493	9.530	0.065
	LMS	-0.061	1.132	0.477	10.382	0.076
	ES	-0.054	1.050	0.680	10.016	0.056
DB return		-0.002	0.019	-0.293	7.121	0.123
$\hat{\varepsilon}_t$	SSA	-0.097	1.142	-0.661	7.868	0.317
	LMS	-0.100	1.132	-0.631	8.855	0.308
	ES	-0.087	1.025	-0.558	6.561	0.242
EURUSD return		0.000	0.006	-0.277	3.903	0.276
$\hat{\varepsilon}_t$	SSA	-0.008	1.091	-0.172	4.190	0.317
	LMS	-0.006	1.074	-0.051	4.175	0.258
	ES	-0.014	1.043	-0.278	3.773	0.270
EURJPY return		0.000	0.008	0.237	6.399	0.275
$\hat{\varepsilon}_t$	SSA	-0.007	1.121	0.164	4.942	0.313
	LMS	-0.006	1.092	0.186	4.953	0.274
	ES	-0.010	1.051	0.164	4.646	0.292

Table 4.6: Descriptive statistics of the real data. The critical value of the KPSS test without trend is 0.347 (90%).

Minimum regulatory requirement: According to the Basel Accord (1998), banks are subject to a credit risk charge and allowed to use an internal model to measure their market risks. To evaluate the internal model, a "traffic light" rule was introduced. It classifies the model into green, yellow or red zone if there are respectively $0-4$, $5-9$ or more than 9 exceptions over Value-at-Risk (VaR) at 1% probabilty level spanning the last 250 days, where

$$\mathrm{VaR}_{t,1\%} = -\sqrt{\hat{\theta}_t} * \mathrm{quantile}(\varepsilon_t)_{1\%}$$

The internal model located in the green zone is accepted. This requirement actually suggests banks to control their VaR at $1.6\%(\frac{4}{250})$ level instead of 1%. Therefore, for fulfilling the regulatory requirement, a model is preferred by giving an empirical probability that is close and not larger than 1.6%, and simultaneously asking for a small amount of risk charge. In the comparison, the sum of VaR over the time horizon is considered as the risk charge:

$$\text{Risk charge} = \sum_t \mathrm{VaR}_t.$$

Table 4.7 gives a detailed report of the calculation. In the table, it shows that all the considered models locate either in the green or yellow zone. Among them, the normal distribution is preferred by presenting the ideal probability and small amount of risk charges, although it less informs the tail-behavior of the underlying series. Furthermore the LMS (for MC, VW and EURUSD) and the SSA (for DB) generate the most desirable results based on the minimum regulatory requirement. The EURJPY data is extraordinary by which the models with the normal noise can not fulfill the regulatory requirement. A compensate choice is the ES with the NIG noise.

Investors' review: From the viewpoint of investors, it is interesting to measure the size of loss instead of the frequency of loss. Since investors suffer loss once the risk controlling is failed and even in the "best" case, the loss equals to the difference between the total realized loss and the reserved risk capital. Therefore, the expected shortfall (ES) has been considered as a better (or coherent) risk measure than the VaR.

$$\mathrm{ES} = \mathsf{E}\{-R_t | - R_t > \mathrm{VaR}_t\}.$$

It is clear that investors will suffer less from a bankruptcy if the amount of losses is smaller. The table shows that the model with normal innovations fails to gauge the market risk, which present smaller empirical probabilities than expected, meaning a high risk and the Gaussian-based model generates larger values of ES than the NIG-based model. Among the NIG models, the

LMS and SSA are desirable. The ES values of EURJPY at the expected 0.5% level, for example, are 0.231 (SSA), 0.255 (LMS) and 0.263 (ES) with NIG innovations. Therefore the combination of the SSA and the NIG combination is desired than the other two.

Internal supervisory review: For internal risk analyzing and controlling, it is desirable to choose a model that can accurately predict the target risk level. Based on this criterion, the model with the NIG innovations has a much better predictability than the others. The table shows once again that the local exponential smoothing generates the most precise values at the two risk levels.

On summary, the calculations based on the local volatility estimates and the NIG distributed residuals best suit the tastes of investors and supervisory. The VaR model based on the local volatilities and the normal distributional assumption, on the contrary, is successful to fulfill the regulatory requirement.

	1% $\varepsilon_t \sim$ N(0, 1)			1% $\varepsilon_t \sim$ NIG			0.5% $\varepsilon_t \sim$ N(0, 1)			0.5% $\varepsilon_t \sim$ NIG		
MC	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES
except.	12	11	8	7	6	5	10	7	6	4	3	3
prob.	0.018	0.016	0.012	0.010 ^{s}	0.009	0.007	0.015	0.010	0.009	0.006	0.004 ^{s}	0.004
$\sum$ ES	0.409	0.377	0.325	0.317	0.285	0.265 ^{i}	0.374	0.303	0.286	0.225	0.193 ^{i}	0.202
$\sum$ VaR	17.18	17.43 ^{r}	18.92	22.08	21.94	21.94						
VW	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES
except.	12	10	10	7	8	6	8	8	7	2	2	3
prob.	0.018	0.015	0.015	0.010 ^{s}	0.012	0.009	0.012	0.012	0.010	0.003	0.003	0.004 ^{s}
$\sum$ ES	0.623	0.567	0.567	0.443	0.492	0.360 ^{i}	0.488	0.488	0.439	0.167 ^{i}	0.167	0.215
$\sum$ VaR	27.83	28.37 ^{r}	28.82	32.44	32.58	33.21						
DB	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES
except.	10	10	7	5	4	6	8	7	4	3	3	3
prob.	0.015	0.015	0.010	0.007	0.006	0.009 ^{s}	0.012	0.010	0.006	0.004 ^{s}	0.004	0.004
$\sum$ ES	0.397	0.397	0.285	0.190	0.148 ^{i}	0.259	0.323	0.301	0.168	0.099 ^{i}	0.099	0.126
$\sum$ VaR	28.00 ^{r}	28.35	29.06	31.23	31.84	30.19						
EURUSD	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES
except.	34	30	22	15	16	18	20	21	9	11	10	7
prob.	0.017	0.015	0.011	0.008	0.008	0.009 ^{s}	0.010	0.010	0.004	0.005	0.005 ^{s}	0.003
$\sum$ ES	0.417	0.372	0.309	0.207 ^{i}	0.212	0.248	0.255	0.254	0.134	0.149	0.143	0.105 ^{i}
$\sum$ VaR	28.00	28.35 ^{r}	28.65	29.51	29.95	29.53						
EURJPY	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES	SSA	LMS	ES
except.	52	50	41	21	20	21	34	30	28	10	11	10
prob.	0.026	0.025	0.020	0.010	0.010 ^{s}	0.010	0.017	0.015	0.014	0.005 ^{s}	0.005	0.005
$\sum$ ES	0.884	0.900	0.797	0.442	0.428 ^{i}	0.463	0.655	0.597	0.572	0.231 ^{i}	0.255	0.263
$\sum$ VaR	32.53	33.09	33.67	40.32	40.21 ^{r}	40.31						

Table 4.7: Risk analysis of the real data. The exceptions are marked in green, yellow or red according to the traffic light rule. An internal model is accepted if it is in the green zone. The best results to fulfill the regulatory requirement are marked by ^{r}. The recommended method to the investor is marked by ^{i}. For the internal supervisory, we recommend the method marked by ^{s}.

Part III

Adaptive Risk Management - Multivariate Models

Chapter 5

Adaptive risk management 3: ICVaR

5.1 Introduction

Risk measurement of large portfolios is a challenging task - both numerically and statistically. The popular risk measure, VaR for example indicates the possible loss over a given time horizon at a risk level pr. From a statistical point of view, it is the pr-quantile of the joint distribution of the portfolio's risk factors which are modelled as:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t), \tag{5.1}$$

where $x(t) \in \mathbb{R}^d$ is the risk factor vector, e.g. the (log) returns of d individual financial instruments. The matrix $\Sigma_x(t)$ denotes the corresponding time-dependent covariance and $\varepsilon_x(t)$ is the d-dimensional standardized residual vector. The portfolio VaR calculation becomes technically difficult for high-dimensionality of the portfolio.

In order to solve this and other numerical problems, portfolio variations are typically mapped into a conditional multivariate Gaussian framework such as the RiskMetrics launched by J.P. Morgan. Recall that, Gaussian distributed residuals or standardized returns $\varepsilon_x(t) = \Sigma_x^{-1/2}(t)x(t)$ are independent and hence the joint density of the residuals is the product of d marginals. In this sense, the portfolio VaR calculation is simplified and only covariance based. Under the Gaussian distributional assumption, many covariance estimation methodologies in a high-dimensional space have been well developed and applied in practice, for example, the constant conditional correlation (CCC) model proposed by [Bol90] and the subsequent dynamic conditional correlation (DCC) model proposed by [Eng02], [ES01]. The simplicity of this kind of covariance based methodology nevertheless bears a risk

of modelling bias since, among other things, the assumed conditional Gaussian marginals are unable to mimic the heavy tailedness of financial time series observed in markets. This issue has been addressed in a variety of papers. For example, [JJ02] have studied the conditional Gaussian distribution fits to VaR that delivers satisfactory estimate at a moderate (e.g. 5%) risk level but underestimates VaR at more extreme level such as 1%. A further example has been illustrated in the real data analysis, see Chapter 3.

The tail problems are evident in Figure 5.1, where we compare the marginal density estimations of the standardized returns of foreign exchange rates, the German Mark to the US Dollar (DEM/USD), from 1979/12/01 to 1994/04/01. In order to mimic the empirical distributional behavior of the real data, we assume 3 different distributional types. The HS-RM (RiskMetrics) and HS-ES$t(15)$ methods fit the standardized returns, based on a GARCH(1,1) volatility process: $\hat{\sigma}_t^2 = 1.65 * 10^{-6} + 0.07x_{t-1}^2 + 0.89\hat{\sigma}_{t-1}^2$, by the Gaussian and Student-t with degrees of freedom 15 distributions, see Chapter 2. The GHADA technique assumes that the standardized returns follow a time homogeneous generalized hyperbolic (GH) distributional mechanism based on locally constant volatilities, see [CHJ05] and [MS04b] for details. The nonparametric kernel density estimation corresponding to the standardized return processes is considered as benchmark. According to the graphical comparison, the GHADA technique is superior to the other two techniques since the empirical GH density coincides to the benchmark, especially in the tails. The HS-ES$t(15)$ technique shows although a better tail fit compared to the RiskMetrics, it still deviates from the benchmark. This small comparison provides evidence that the Gaussian distributional assumption is unreliable and will lead to low accuracy of univariate and portfolio VaR calculations.

The weak performance of the Gaussian assumption motivates us to search for a different approach solving the technical problems of portfolio VaR calculations. An "ideal" situation is, as mentioned before, that the residuals $\varepsilon_x(t)$ are independent. Since based on the independence, the estimation of the joint distribution can be converted to marginals' estimations.

In the context of sound engineering, signal detection from unknown filters and sources is treated by a method called independent component analysis (ICA). This engineering method is designed for detection of blind folded signals and retrieves out of a high-dimensional time series stochastically independent source components. A tutorial on ICA can be found in [HO99] and a variety of numerical techniques to uncover independent components (ICs) are discussed in [HKO01]. Promoted by the success in engineering, ICA has been applied in different areas such as brain imaging [DJK$^+$02] and telecommunication study [RRK02]. An early implementation of ICA in financial time series is given in [BW98], drawing comparisons of ICs and principal

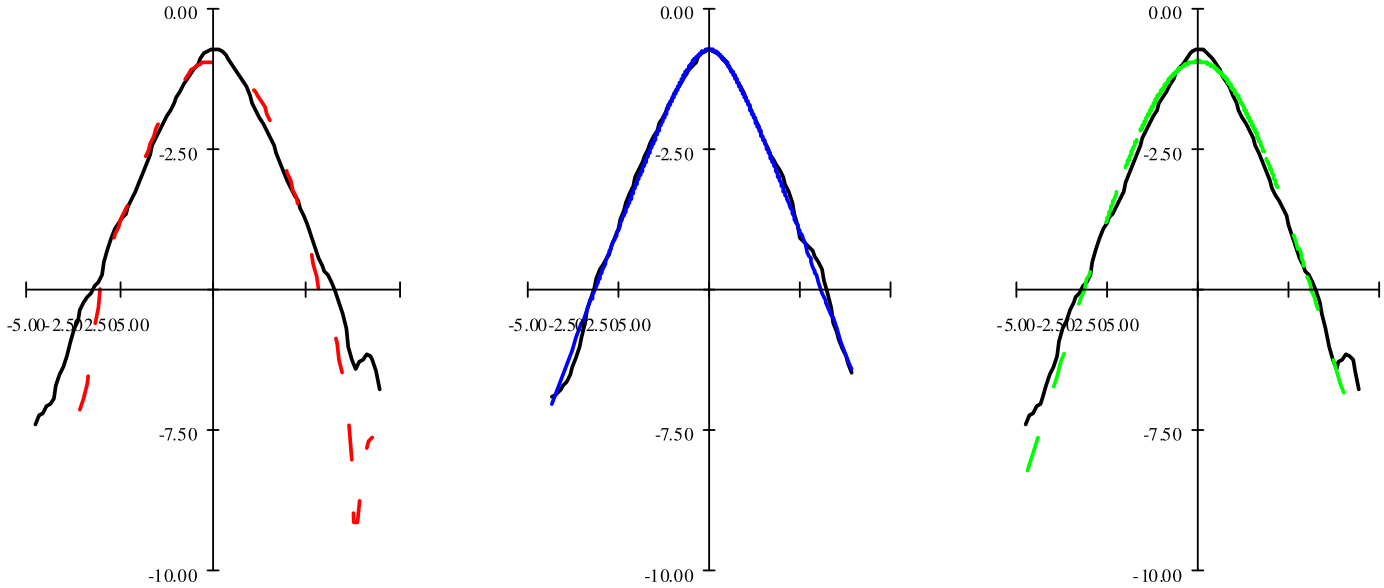

Figure 5.1: Graphical comparison of density estimations based on the standardized DEM/USD returns from 1979/12/01 to 1994/04/01 (3719 observations). The nonparametric kernel density estimations (solid curve) are considered as benchmarks. Notice that the nonparametric density estimations result distinct forms w.r.t. different standardization, or more details, w.r.t. different volatility estimation methods. The HS-RM based density estimation is the dotted curve on the left panel whereas the HS-ES$t(15)$ based estimation is displayed on the right. The corresponding GARCH(1,1) process is: $\hat{\sigma}_x^2(t) = 1.65 * 10^{-6} + 0.07x^2(t-1) + 0.89\hat{\sigma}_x^2(t-1)$. The GHADA technique with the local constant volatility process and the GH$(1, 1.74, -0.02, 0.78, 0.01)$ density estimation is displayed on the middle panel. Data source: FEDC(http://sfb649.wiwi.hu-berlin.de).

components (PCs) applied to 28 Japanese stocks from 1986 to 1989. Few contributions however exist for the application of ICA in risk management.

The first aim of this chapter is to bring together the lines of thought of the engineering signal processing literature and newer statistical insights on the high-dimensional VaR calculations. We coin the name ICVaR owing to the ICA technology. Notice that the technique is as well applied in the calculation of expected shortfall (ES). Given a trading strategy $b(t) \in \mathbb{R}^d$, the portfolio return $r(t) \in \mathbb{R}$ is:

$$r(t) = b(t)^\top x(t). \tag{5.2}$$

The ICVaR technique does not rely on a direct joint density estimation of the high-dimensional returns $x(t) \in \mathbb{R}^d$ such as $x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t)$. Instead the ICVaR procedure in a first step applies a linear transformation to $x(t)$, namely a nonsingular matrix W yields (approximately) ICs $y(t) \in \mathbb{R}^d$:

$$x(t) = W^{-1}y(t) \tag{5.3}$$

The matrix W is different from the Mahalanobis transformation $\mathrm{cov}(x)^{-1/2}$ - that creates ICs in the Gaussian regime - only in the case of non-Gaussian marginals as we will see later. The second ICVaR step concerns the fit of each IC univariately:

$$\begin{aligned} y(t) &= \mathrm{diag}\{\sigma_{y_1}(t), \cdots, \sigma_{y_d}(t)\}\varepsilon_y(t) = D_y^{1/2}(t)\varepsilon_y(t) \\ \text{or equivalently: } y_j(t) &= \sigma_{y_j}(t)\varepsilon_{y_j(t)}, \quad j = 1, \cdots, d, \end{aligned} \tag{5.4}$$

where the covariance $D_y(t)$ of the ICs is diagonal due to independence and $\varepsilon_y(t)$ are cross independent innovations. Based on the ICA, the high-dimensional VaR problem is now converted to simpler univariate VaR calculations. Given the discussion above on fitting VaR this opens a wide avenue of alternative VaR determination.

The second purpose of this chapter is to compare the proposed ICVaR technique with the industry standard RiskMetrics and the most-often used HS-ES$t(df)$ method. As discussed before, the HS-RM often gives underestimated VaRs, an inevitable cost of the Gaussian assumption. On the contrary, the Students-t distribution can better mimic the heavy-tailed distributional behavior of financial risk factor than the HS-RM, but it is hard to reflect the leptokurtic scenario exactly, as illustrated in Figure 5.1. In this chapter, we show how high accuracy can be reached by the ICVaR compared to the HS-RM and HS-ES$t(df)$ methods in real data analysis. Since the distributional type of the portfolio returns $r(t) = b(t)^\top x(t)$ is not clear under many realistic distributional assumption, such as the Student-t, the Monte Carlo simulation method is applied in this case, see Chapter 2.

The methodological contribution of the study unfolds in Section 2 where the ICA method is discussed. The simulation study is presented in Section 3. Further we apply the ICVaR to exchange rate portfolios with different artificial trading strategies. The ICVaR predicts risk levels precisely and outperforms the HS-RM and HS-ES$t(df)$ methods. Finally we conclude our study in Section 5.

5.2 ICVaR methodology

5.2.1 Basic model

The proposed ICVaR methodology consists of 2 main steps: searching for ICs based on a linear transformation and modelling the univariate volatility and stochastic distribution.

$$\begin{aligned} r(t) &= b(t)^\top x(t) \\ &= b(t)^\top W^{-1} y(t) \\ &= b(t)^\top W^{-1} D_y^{1/2}(t) \varepsilon_y(t). \end{aligned}$$

The idea of the ICA is that risk factors $x(t) \in \mathbb{R}^d$ can be represented by a linear combination of d-dimensional ICs. The linear transformation matrix W is assumed to be nonsingular. Due to the independence property of ICs, the covariance $D_y(t)$ must be a diagonal matrix and the elements of the stochastic vector $\varepsilon_y(t)$ are cross independent. Furthermore, it fulfills that $\mathsf{E}[\varepsilon_y(t)|\mathcal{F}_{t-1}] = 0$ and $Var[\varepsilon_y(t)|\mathcal{F}_{t-1}] = I_d$.

From a statistical viewpoint, this projection technique is desirable since the d-dimensional portfolio is decomposed to univariate and independent risk factors through a simple linear transformation. Recall that, the joint density (f) and the covariance of any linear transformed ICs such as $x(t) = W^{-1}y(t)$ are analytically computable:

$$\begin{aligned} f_y &= \prod_{j=1}^{d} f_{y_j}, \\ f_x &= \text{abs}(|W|) f_y(Wx), \end{aligned} \qquad \begin{aligned} \Sigma_y(t) &= D_y(t) \\ \Sigma_x(t) &= W^{-1} D_y(t) W^{-1\top}. \end{aligned}$$

In the second step of the proposed ICVaR, the diagonal elements of $D_y(t)$ and each univariate component of $\varepsilon_y(t)$ are estimated since the matrix manipulation is equivalent to:

$$y_j(t) = \sigma_{y_j}(t)\varepsilon_{y_j}(t), \quad j = 1, \cdots, d, \tag{5.5}$$

where $\sigma_{y_j}(t)$ is the square root of the j-th diagonal element of $D_y(t)$ and $\varepsilon_{y_j}(t)$ is the univariate stochastic term with $\mathsf{E}[\varepsilon_{y_j}(t)|\mathcal{F}_{t-1}] = 0$ and $Var[\varepsilon_{y_j}(t)|\mathcal{F}_{t-1}] =$

1. There are various univariate models available to estimate the volatility and approximate the distribution of the stochastic term. For the fit of the marginal IC factors we refer to the GHADA technique as in [CHJ05], where one adaptively specifies the local smoothing interval, by which the volatility is estimated using the average squared returns. The standardized returns are identified in the GH distributional framework. An alternative approach is given by the HS-ES$t(df)$ setup. It could be used to estimate the heteroscedastic volatility process and the Student-t could be applied to pick up the heavy-tailedness of ε_j.

After these two steps, one may compute the quantile of portfolio risk by Monte Carlo (MC) simulation or analytical methods e.g. saddle point approximation, see [IM98]. For the ease of presentation, we concentrate in this chapter on the simulation methodology. To be more specific, we generate $d-$dimensional samples of the fitted distributions with sample size M, from which we calculate the daily empirical pr-quantile of the portfolio variations. The simulation will repeat N times and the average value of the empirical quantiles is considered as the portfolio VaR at level pr:

$$\mathrm{VaR}_{\mathrm{pr},t} = \frac{1}{N}\sum_{n=1}^{N} \hat{F}^{-1}_{\mathrm{pr},t}\{r^{(n)}(t)\} = \frac{1}{N}\sum_{n=1}^{N} \hat{F}^{-1}_{\mathrm{pr},t}\{b(t)^\top \hat{W}^{-1}\hat{D}_y(t)^{1/2}\hat{\varepsilon}_y^{(n)}(t)\},$$

where $\hat{F}^{-1}_{\mathrm{pr},t}$ denotes the empirical quantile function of $r(t)$.

5.2.2 ICA: Properties and Estimation

Since ICA is a relatively new technique in this context, we present a small pedagogical illustration of its usage.

Example: Generate 3 independent GH random variables GH$(y; 1, 2, 0, 1, 0)$, GH$(y; 1, 1.7, 0, 0.5, 0)$ and GH$(y; 1, 1.5, 0, 1, 0)$ as sources. The first distributional parameter specifies the subclass of the random variables. With the value of 1, they are hyperbolic (HYP) distributed. The other four parameters control the location, scale, asymmetry and likeliness of extreme events, see Chapter 3. These source components have mean $(-0.02, 0.05, -0.00)^\top$ and standard deviation (sd) $(0.83, 0.92, 0.99)^\top$ respectively. The linear transformation matrix is the estimate based on 3 real German stocks' returns: Allianz, BASF and Bayer from 1974/01/02 to 1996/12/30 with value of:

$$W^{-1} = \begin{pmatrix} 1.31 & 0.14 & 0.18 \\ -0.42 & -1.26 & -1.25 \\ -0.03 & 0.41 & -0.49 \end{pmatrix} 10^{-2} \tag{5.6}$$

The generating time series $x(t) = W^{-1}y(t)$ are analyzed by the ICA. The covariance of $x(t)$ is

$$\Sigma_x = \begin{pmatrix} 13.66 & 6.04 & 6.49 \\ 6.04 & 15.54 & 11.64 \\ 6.49 & 11.64 & 16.07 \end{pmatrix} 10^{-5}, \tag{5.7}$$

which is identical to $W^{-1}D_y(t)W^{-1\top}$. Recall that in the Gaussian framework, the Mahalanobis transformation delivers independent variables:

$$\widehat{\Sigma}_x^{-1/2} = \begin{pmatrix} 0.91 & -0.09 & -0.12 \\ -0.09 & 1.03 & -0.41 \\ -0.12 & -0.41 & 1.04 \end{pmatrix} 10^2 \tag{5.8}$$

which is clearly distinct from

$$W = \begin{pmatrix} 0.79 & 0.10 & 0.03 \\ -0.11 & -0.44 & 1.08 \\ -0.15 & -0.38 & -1.10 \end{pmatrix} 10^2 \tag{5.9}$$

indicated by (5.6). Figure 5.2 provides an illustration of this procedure. The top row contains the 3 independent source signals $y(t)$. The middle row displays the time series $x(t) = W^{-1}y(t)$. One sees the scale changes and different random patterns display. The last row of Figure 5.2 shows the ICs estimated by the ICA method. The time series on the top and the bottom look familiar but the sign and the ordering may change as displayed in the figure. The first estimated IC for example displays the similar movement as the third generated IC. Furthermore, the third estimated IC has a mirror pattern of the second true IC. Notice that the estimated linear transformation matrix W has corresponding change in sign. Therefore, this sign-identification problem does not influence the density estimation of $x(t) = W^{-1}y(t)$.

Scale identification

In fact, the scales of $y(t)$ and W are not identifiable. Given a matrix $C = \text{diag}(c_1, \cdots, c_d)$ with $c_j \neq 0$, for example, the new ICs (Cy_t) with the transformation $(W^{-1}C^{-1})$ also fulfill (5.3). In order to avoid the identification problem, it is suggested to prewhiten $x(t)$ and assume $y(t)$ to be standardized. The Mahalanobis transformation $\widehat{\Sigma}_x^{-1/2}$ does the prewhitening job. It is not hard to see that W becomes then an orthogonal matrix. Denote by $\tilde{x}_t$ - the prewhitened $x(t)$ and $y(t) = \tilde{W}\tilde{x}(t)$ the corresponding ICs. $W = \tilde{W}\widehat{\text{cov}}(x)^{-1/2}$ is then the linear transformation for the original observations. For notational convenience, we assume from now on that $x(t)$ has been prewhitened.

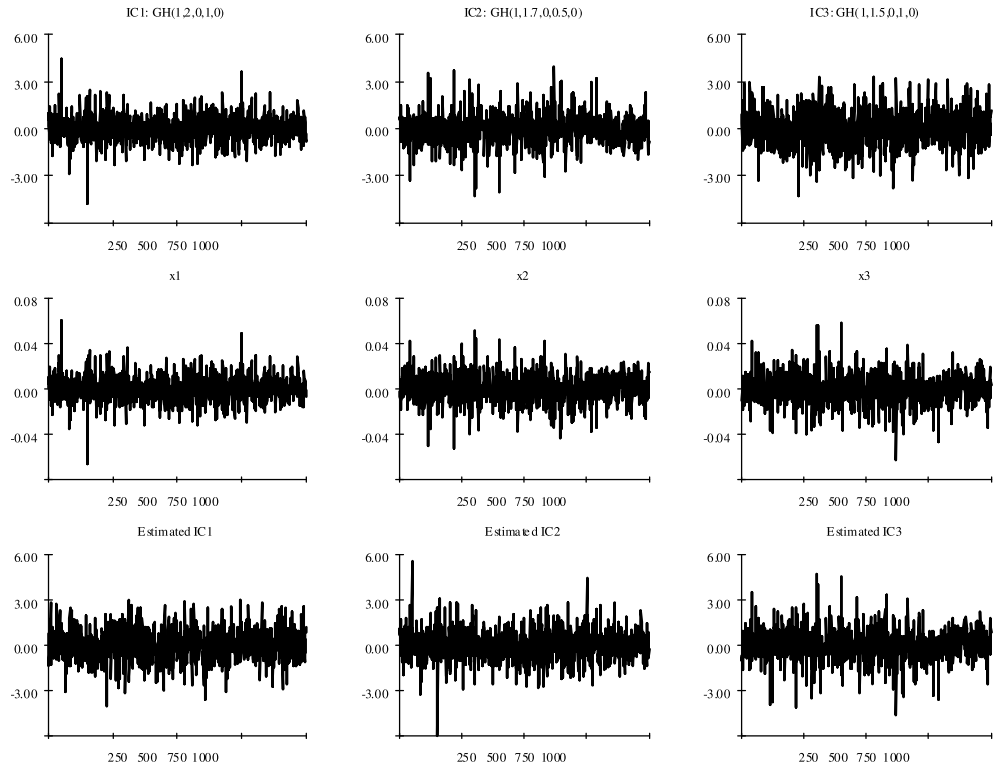

Figure 5.2: Simulation: on the top are time plots of 3 independent GH variables. The mixed variables $x(t) = W^{-1}y(t)$ are displayed on the middle row where the linear transformation matrix W is estimated based on 3 German stocks' returns: Allianz, BASF and Bayer from 1974/01/02 to 1996/12/30. The time series of the estimated ICs are displayed on the bottom.

Order identification
Furthermore, the order of the ICs is ambiguous. Given a permutation matrix P, the ICs (Py_t) fulfill (5.3) with a new transformation $(W^{-1}P^{-1})$ as well.
IC is necessarily non-Gaussian
Given a d-dimensional standardized Gaussian vector $x(t)$ and an orthogonal matrix W, the joint pdfs of $x(t)$ and $y(t)$ are in fact identical and unrelated with W:

$$\begin{aligned} f_x &= |2\pi \mathrm{I}_d|^{-\frac{1}{2}} \exp\left(-\frac{x^\top x}{2}\right) \\ &= f_x(W^{-1}y)|\det W^{-1}| \\ &= |2\pi \mathrm{I}_d|^{-\frac{1}{2}} \exp\left(-\frac{y^\top y}{2}\right) = f_y. \end{aligned}$$

This condition seems strict, but it is naturally fulfilled in financial applications, where financial time series, even after standardization, display heavy-tailed behavior, see [ABDL01].

How can we estimate the linear transformation matrix W after we have prewhitened the non-Gaussian variables? Independence of the components of a random vector $y \in \mathbb{R}^d$ can be measured by the mutual information, see [VW97]:

$$\begin{aligned} I(W,y) &= I(W,f_y) = \sum_{j=1}^{d} H(y_j) - H(y) \\ &= \sum_{j=1}^{d} H(y_j) - H(x) - \log|\det(W)| \end{aligned} \tag{5.10}$$

where $H(y) = H(f_y) = -\int f_y(u)\log f_y(u)du$ is the entropy of the vector y with a joint pdf f_y. If the components of y are independent, then the mutual information will reach its minimum with a value of 0. Therefore, the IC searching is identical to minimizing (5.10) w.r.t. W. Since $H(x)$ is fixed given the data, and the matrix W is orthogonal after prewhitening, this problem is further equivalent to minimizing the term $\sum_{j=1}^{d} H(y_j)$. Now we replace the objective function of the optimization problem by $\min H(y_j)$ w.r.t. w_j, the j-th row of W. Notice that:

$$\sum_{j=1}^{d} \min H(y_j) \;\le\; \min \sum_{j=1}^{d} H(y_j) \;=\; \min \sum_{j=1}^{d} H(w_j x) \tag{5.11}$$

This replacement leads to some loss in the W estimation but extensively speeds up the estimation procedure. Moreover, the entropy and negentropy

$J(w_j, y_j) = H\{\mathrm{N}(0,1)\} - H(y_j)$ are in one-to-one correspondence, we can also formulate the optimization problem as:

$$\hat{w}_j = \mathrm{argmin} H(y_j) = \mathrm{argmax} J(w_j, y_j).$$

The negentropy is always nonnegative, since with a fixed variance, the Gaussian random variable has the largest entropy among all distributional types, [CT91]. Therefore, the negentropy is considered as a non-Gaussian measure and widely used in projection pursuit (PP), see [JS87]. In this sense, the PP methods of searching non-Gaussian direction can be applied in the IC identification as well. On the other hand, compared to the cumulant based PP method, the entropy or negentropy is less sensitive to outliers and therefore preferable.

It arises question at this stage, i.e. the marginal pdfs of the ICs in the entropy or the negentropy are unknown. A distributional free approximation of the univariate negentropy has been proposed by [Hyv98]:

$$J(w_j, y_j) \approx C \left\{\mathsf{E}\{G(y_j)\} - \mathsf{E}[G\{\mathrm{N}(0,1)\}]\right\}^2 \tag{5.12}$$

where C is a constant and G is an even function, e.g. $G(y_j) = \frac{1}{\kappa}\mathrm{logcosh}(\kappa y_j)$, $1 \leq \kappa \leq 2$.

Excursion of the negentropy approximation in (5.12): The motivation is to maximize the univariate negentropy under the "worst" situation, where the interesting random variable belongs to the density family that gives us the minimum negentropy:

$$f_0(y; a) = A \exp\{\sum_{s=1}^{\infty} a_s G_s(y)\} \tag{5.13}$$

where $G_s(\cdot)$ is function whose expectation $\mathsf{E}[G_s(y)]$ is used to approximate f_y. By doing so, the density approximation is expected to give us a really maximized negentropy on general. Normally two G functions are used, where G_1 is an odd function and G_2 is even. The univariate negentropy approximation can be formulated as:

$$J(w_j, y_j) \approx k_1[\mathsf{E}\{G_1(y_j)\}]^2 + k_2[\mathsf{E}\{G_2(y_j)\} - \mathsf{E}\{G_2(y_{gauss})\}]^2 \tag{5.14}$$

where k_1 and k_2 are positive constants in accordance with different functions G_s. For example, by choosing:

$$G_1^a(y_j) = y \exp(-y_j^2/2) \tag{5.15}$$

$$G_2^a(y_j) = \exp(-y_j^2/2) \tag{5.16}$$

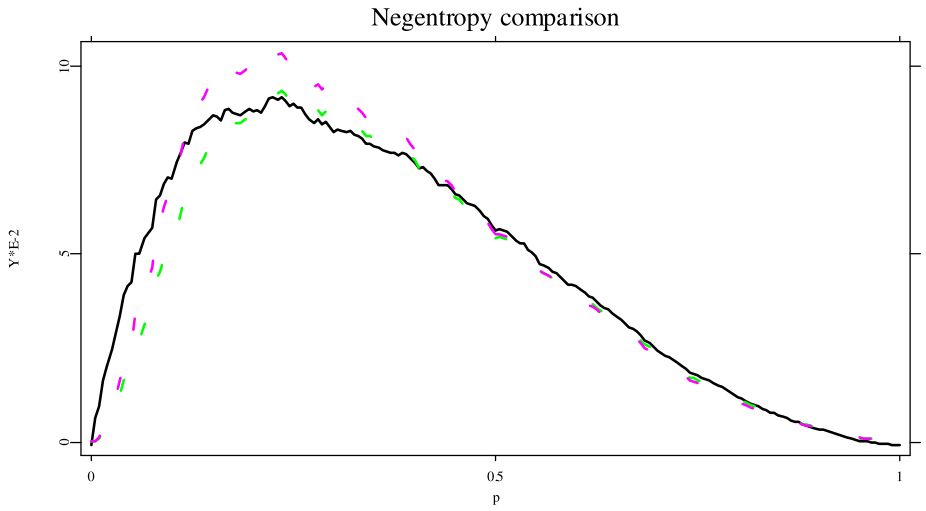

Figure 5.3: Comparison of the true negentropy (solid line) and its approximations (a: red and dashed, b: blue and dotted) of a simulated Gaussian mixture variable: $p\mathrm{N}(0,1) + (1-p)\mathrm{N}(1,4)$ for $p \in [0,1]$.

we have $k_1 = 36/(8\sqrt{3}-9)$ and $k_2^a = 1/(2-6/\pi)$. Using another approximation by which

$$G_1^b(y_j) = y_j \exp(-y_j^2/2) \tag{5.17}$$
$$G_2^b(y_j) = |y_j|, \tag{5.18}$$

we have $k_1 = 36/(8\sqrt{3}-9)$ and $k_2^b = 24/(16\sqrt{3}-27)$.

Figure 5.3 compares the true negentropy of a simulated mixture variable $y = p\mathrm{N}(0,1)+(1-p)\mathrm{N}(1,4)$ where $p \in [0,1]$ and $\mathrm{N}(m,v)$ denotes a univariate Gaussian variable with a mean of m and a variance of v. Given different value of p, the two proposed negentropy approximations are very close to the true negentropy. It states that these approximations are at least numerically reliable.

A further simplification is to assume the symmetry of ICs, where the odd function G_1 disappears.

The estimation of the linear transformation matrix in the ICA is estimated by:

$$\hat{w}_j = \operatorname{argmax} \{\mathsf{E}\{G(w_j X)\} - \mathsf{E}[G\{\mathrm{N}(0,1)\}]\}^2 \tag{5.19}$$
$$s.t. \qquad W^\top W = \mathrm{I}_d \tag{5.20}$$

Based on the Kuhn-Tucker condition and the Newton's method, [HO99] have proposed the FastICA algorithm.
FastICA algorithm: set $j = 1$

1. Choose an initial vector w_j of unit norm, $W = (w_1, \cdots, w_d)^\top$.
2. Let $w_j^{(n)} = \mathsf{E}\left[g\{w_j^{(n-1)}x(t)\}x(t)\right] - \mathsf{E}\left[g'\{w_j^{(n-1)}x(t)\}\right] w_j^{(n-1)}$, where g denotes the first derivative of $G(y_j)$ and g' the second derivative. In practice, the sample mean is applied for $\mathsf{E}[\cdot]$.
3. Orthogonalization: $w_j^{(n)} = w_j^{(n)} - \sum_{k \neq j}(w_j^{(n)\top} w_k)w_k$.
4. Normalization: $w_j^{(n)} = w_j^{(n)}/||w_j^{(n)}||$, $||\cdot||$ denotes the norm.
5. If the result does not converge, i.e. $||w_j^{(n)} - w_j^{(n-1)}|| \neq 0$, go back to 2.
6. Set $j = j + 1$. For $j \leq d$, go back to step 1.

5.3 Simulation Study

The accuracy of the proposed ICVaR depends on the linear transformation matrix estimation and the univariate modeling on the ICs estimated. In order to fit the local distributions of ICs, we apply the GHADA technique due to its good performance in the simulation and empirical studies in [CHJ05]. The target of the simulation study here is to search for ICs and compare the marginal densities of the estimated and generated ICs.

In particular, we pursue an experiment to check the validation of the FastICA approach with normal-inverse Gaussian (NIG) distributed ICs, where the NIG distribution is a subclass of the GH distribution with the fixed parameter $\lambda = -0.5$, see [BNB81]. We generate $d = 50$ NIG samples with $T = 1000$ observations, i.e. $y_j \sim \text{NIG}(\alpha_j, \beta_j, \delta_j, \mu_j)$ for $j = 1, \cdots, 50$. Without loss of generality, we set $\mu_j = 0$ and $\delta_j = 1$. The parameter α_j is uniformly distributed in $[1, 2]$ and β fulfills the condition:

$$Var[y_j] = \frac{1}{\sqrt{\alpha_j^2 - \beta_j^2}} \frac{\alpha_j^2}{\alpha_j^2 - \beta_j^2} = 1,$$

such that the generated ICs have unit variances. Furthermore, the sign of β is chosen arbitrarily. Table 5.1 shows the distributional parameters of the generated ICs. The linear transformation matrix W^{-1} is obtained via the Jordan decomposition of a square matrix, whose elements are standard

ICest	$\hat{\alpha}$	$\hat{\beta}$	IC	α	β	MAE	ICest	$\hat{\alpha}$	$\hat{\beta}$	IC	α	β	MAE
1	1.89	0.63	23	1.39	0.61	0.63	26	1.84	0.60	2	1.92	1.14	2.56
2	1.76	0.60	25	1.97	1.19	1.57	27	1.82	-0.18	16	1.57	-0.80	2.56
3	1.45	-0.51	30	1.06	-0.21	0.32	28	1.30	-0.37	8	1.17	-0.37	0.32
4	1.64	0.42	36	1.22	0.43	0.52	29	1.34	0.13	17	1.06	0.21	0.68
5	2.11	-0.67	13	1.31	-0.53	0.70	30	2.02	-0.39	44	1.54	0.77	2.73
6	2.05	-0.62	28	1.39	-0.61	0.93	31	3.05	0.80	35	1.78	1.00	3.27
7	1.73	-0.41	50	1.60	-0.83	0.76	32	1.79	-0.45	32	1.34	-0.57	0.77
8	1.62	0.45	7	1.15	0.35	0.45	33	1.58	0.08	4	1.24	0.46	0.88
9	1.27	0.16	6	1.96	1.17	1.58	34	1.59	0.35	47	1.45	0.67	0.74
10	1.51	0.43	27	1.51	-0.74	2.33	35	2.00	-0.33	34	1.43	-0.66	1.18
11	2.44	1.33	38	1.27	0.49	0.98	36	1.63	-0.06	31	1.85	1.07	1.37
12	1.65	-0.25	26	1.35	-0.57	0.79	37	1.81	0.29	29	1.93	1.15	1.29
13	2.58	-1.35	10	1.89	-1.11	1.69	38	1.49	-0.01	22	1.66	-0.89	2.97
14	1.60	0.15	24	1.54	0.77	0.98	39	2.39	-0.30	42	1.53	-0.76	1.27
15	1.52	-0.03	20	1.04	0.18	0.67	40	1.52	-0.22	21	1.39	0.62	1.48
16	2.14	-1.00	1	1.72	-0.94	3.09	41	2.32	-0.001	41	1.70	0.93	3.37
17	2.20	0.64	3	1.75	0.97	4.07	42	2.87	0.30	9	1.50	0.73	1.37
18	1.44	0.48	15	1.40	0.63	2.28	43	2.27	0.24	14	1.28	0.50	1.13
19	1.44	-0.40	33	1.70	0.93	3.22	44	2.44	0.46	43	1.87	1.09	1.39
20	1.90	-0.54	39	1.72	0.95	9.73	45	1.88	-0.07	18	1.44	-0.67	1.23
21	1.57	0.39	11	1.63	0.86	0.72	46	2.22	0.13	37	1.39	0.62	1.24
22	1.80	-0.56	12	1.78	-1.00	0.84	47	1.79	0.26	49	1.45	-0.68	1.41
23	1.69	0.20	46	1.69	0.92	1.09	48	3.03	0.75	40	1.95	1.17	2.37
24	1.76	-0.27	5	1.39	0.62	1.53	49	3.31	-0.18	48	1.50	-0.73	1.49
25	1.50	0.26	19	1.69	-0.92	1.38	50	3.77	0.29	45	1.53	0.76	1.52

Table 5.1: ML estimators $\hat{\alpha}_j$ and $\hat{\beta}_j$ of the estimated ICs, the parameters of the true ICs and the MAE (unit: 10^{-2}).

normally distributed. The mixed time series $x(t) = W^{-1}y(t)$ are analyzed by the ICA.

We apply the FastICA algorithm to the transformed time series $x(t)$ and estimate the NIG parameters of each estimated IC. We order the 50 estimated independent series by minimizing the mean absolute error (MAE) of the marginal pdfs between the estimated and generated ICs:

$$\mathrm{MAE}_j = \frac{1}{T} \sum_{t=1}^{T=1000} |f(\hat{y}_j(t)) - f(y_j(t))|.$$

Overlapping is avoided in the ordering. The largest two MAEs are 0.09 (ICest 20) and 0.04 (ICest 17), indicating the worst cases of IC searching. In this case, it is expected to get accurate VaR estimations based on these fits.

time series	mean	sd	skewness	kurtosis	ρ_1	ρ_2
DEM/USD	0.00	$0.71*10^{-2}$	-0.13	4.94	0.02	0.01
GBP/USD	0.00	$0.69*10^{-2}$	-0.01	5.64	0.08	0.01
IC1	-0.02	1.00	-0.62	8.71	0.07	0.02
IC2	0.01	1.00	-0.08	5.19	0.05	0.01

Table 5.2: Descriptive statistics of the log returns and the two estimated indpendent processes of the DEM/USD and GBP/USD rates.

5.4 Real data analysis

5.4.1 Exchange rate

In this section, we analyze foreign exchange rate portfolios with static trading strategies. The foreign exchange market, or "Forex" market, is by far the largest financial market in the world with trading volumes surpassing USD 1.5 trillion on some days. The very active buying and selling of traders make it further the most liquid financial market. Among others, we consider portfolios including two exchange rates: DEM/USD and the British Pound to the US Dollar (GBP/USD) from 1979/12/01 to 1994/04/01. We forecast VaRs one-day-ahead w.r.t. four artificial trading strategies, i.e. $b_1 = (1,1)^\top$, $b_2 = (1,2)^\top$, $b_3 = (-1,2)^\top$ and $b_4 = (-2,1)^\top$. For example, b_1 means holding one unit DEM/USD and one unit GBP/USD forward contracts over time. Since the position of these individual risk factors are constant in time, one can simply use the historical simulation approach, by considering the portfolio returns as one univariate risk factor, to measure the market risk. It speeds up the VaR computation but may reduce the accuracy of estimation, see Chapter 2. Here we compare three VaR models: the proposed ICVaR approach based on the multivariate risk factors, the HS-RM and HS-ES$t(df)$ methodologies based on the univariate portfolio returns. In the HS-RM and HS-ES$t(df)$ frameworks, we apply the GARCH(1,1) setup to estimate the dependence structure of real data and assume that the standardized returns are Gaussian or Student-t distributed. The degrees of freedom (df) of the Student-t distribution are selected by the maximum likelihood method.

The data is available at FEDC (sfb649.wiwi.hu-berlin.de). Each time series consists of 3721 observations. Table 5.2 summarizes the statistical properties of the original data and the estimated ICs based on the linear transformation. These four time series are all centered around 0, approximately symmetric but obviously non-Gaussian indicated by their large kurtoses. Since the distributional mechanism of the standardized returns is

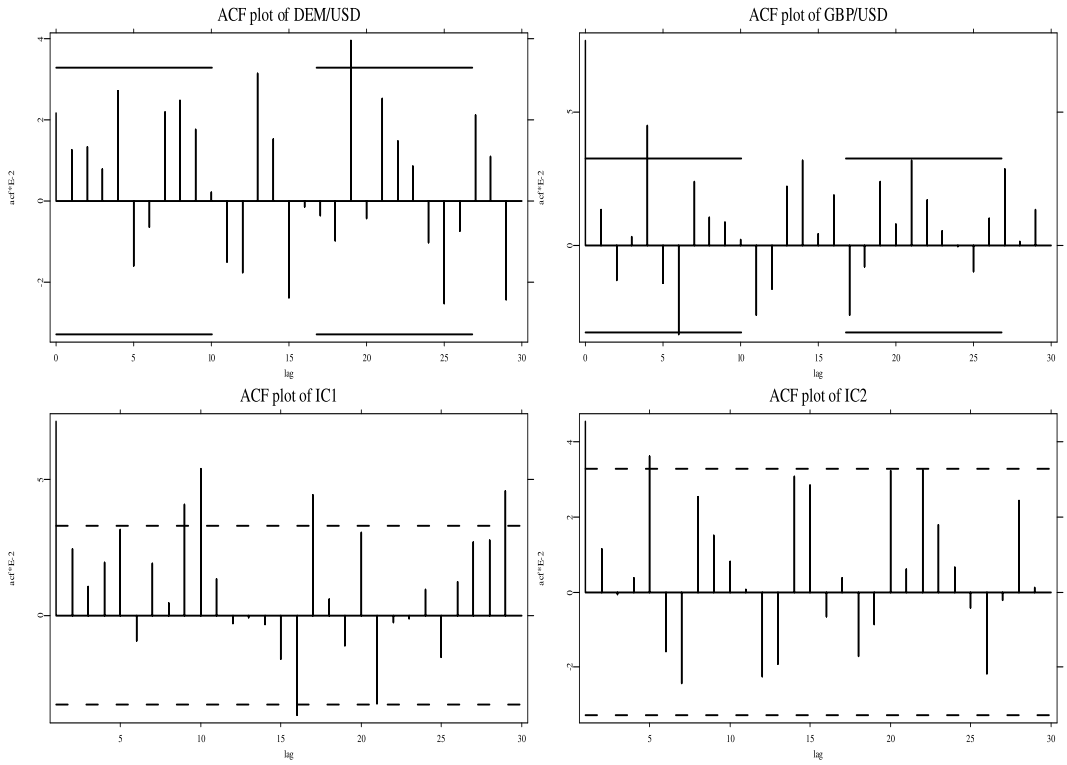

Figure 5.4: ACF plots of the log returns of the DEM/USD (left) and the GBP/USD (right) are displays on the top. Below are the ACF plots of the estimated IC series: IC1 (left) and IC2 (right).

assumed to be time homogeneous, one needs to check the temporal dependence of the series. The ACF plots show that the serial correlations decay at the very beginning lags, indicating a weak stationarity of the underlying series, see Figure 5.4. On the other hand, the cross correlation of the two exchange returns is over 0.77, referring a strongly linear dependence between them.

Applying the FastICA algorithm, we estimate the linear transformation matrix and its inverse:

$$\hat{W} = \begin{pmatrix} 207.93 & -213.63 \\ 77.72 & 73.29 \end{pmatrix}, \quad \hat{W}^{-1} = \begin{pmatrix} 2.30 & 6.71 \\ -2.44 & 6.53 \end{pmatrix} 10^{-3} \tag{5.21}$$

We then implement the GHADA approach to fit the distributional feature of each estimated IC. Figure 5.5 shows the adaptive volatility series based on the two ICs. The well-known volatility clustering is displayed and volatility jumps appear. These jumps happen in most cases at different time and have

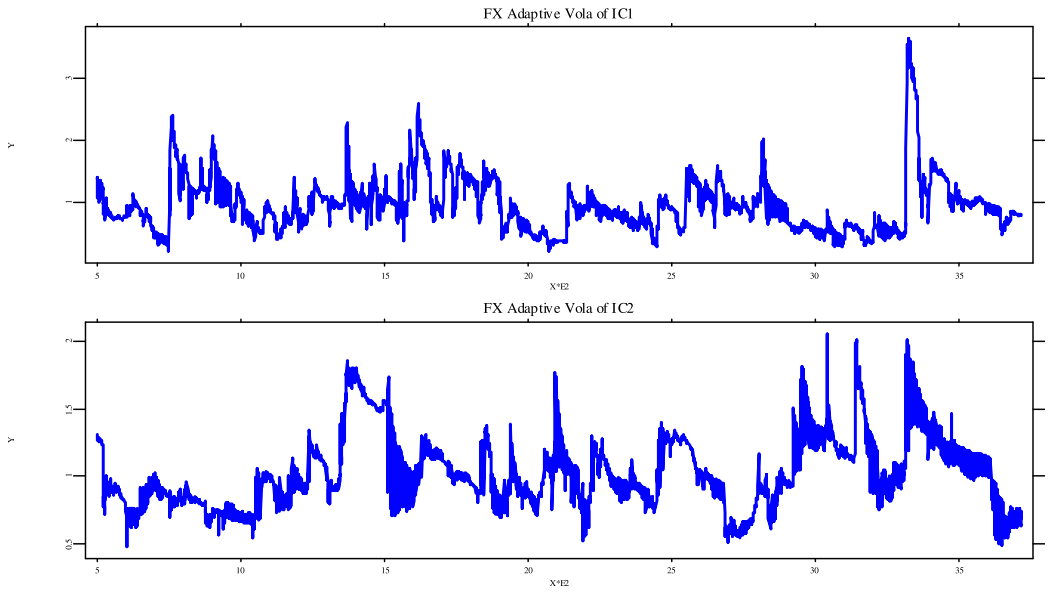

Figure 5.5: Adaptive volatility processes of the FX ICs.

individual influences on the return processes.

The estimated HYP and NIG parameters of the standardized ICs are reported in Table 5.3. To test the independence of the two ICs estimated, we compare the nonparametric joint density and the product of the marginals of the two ICs graphically in Figure 5.6. On the left, the nonparametric joint density of the returns displays a different surface from the bivariate Gaussian fit. On the right, the product of the two marginals displays a similar surface as the empirical joint density with different scale. According to the Jacobian transformation, the pdf of $x(t)$ is:

$$f_x = \text{abs}(|W|) f_y(Wx).$$

This explains the scale difference and supports not only the independence assumption but also the linear assumption of the ICA. In the MC simulation to find the empirical quantiles of the portfolios, we generate $d = 2$ samples with $M = 10,000$ observations. The daily empirical quantiles at 3 risk levels pr $= 5\%$, 1% and 0.5% are the average values of $N = 100$ repetitions. We implement the simulation for the last $T = 1000$ days. The daily means and standard deviations of the 3 empirical quantiles given different trading strategies and two GH subclasses (HYP and NIG) are reported in Table 5.4 to Table 5.7. The largest standard deviation of the daily empirical quantiles is underlined in each category. The values are small due to the large sample

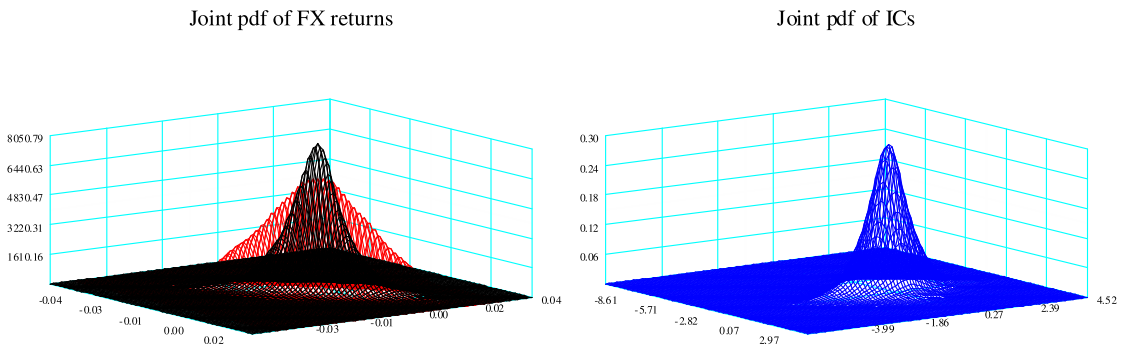

Figure 5.6: Comparison of the nonparametric joint density (black) of the returns of the exchange rates and the product (blue) of the HYP marginal densities of two ICs. The red surface is the Gaussian fitting with the same covariance as the returns of the exchange rates.

GH type	time series	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\mu}$
HYP	IC1	1.71	-0.17	0.55	0.12
HYP	IC2	1.77	0.02	0.71	-0.01
NIG	IC1	1.22	-0.18	1.10	0.13
NIG	IC2	1.37	0.03	1.28	-0.02

Table 5.3: Identified GH parameters of the estimated ICs.

size, indicating efficient estimation of the daily quantiles.

The backtesting results based on the ICVaR, the HS-RM and HS-ES$t(df)$ methodologies are reported in Table 5.8. The degrees of freedom of the Student-t fits are $16, 16, 13$ and 14 w.r.t. different trading strategies. Based on the likelihood ratio tests, LR1 for risk level and LR2 for exceptions clustering [Jor01]:

$$\begin{aligned} \mathrm{LR}_1 &= -2\log\{(1-\mathrm{pr})^{T-N}\mathrm{pr}^N\} + 2\log\{(1-N/T)^{T-N}(N/T)^N\} \\ \mathrm{LR2} &= -2\log\{\hat{\pi}^{n_0}(1-\hat{\pi})^{n_1}\} + 2\log\{\hat{\pi}_{00}^{n_{00}}\hat{\pi}_{01}^{n_{01}}\hat{\pi}_{10}^{n_{10}}\hat{\pi}_{11}^{n_{11}}\} \end{aligned}$$

where N is the number of the exceptions over T time points, $\hat{\pi}_{ij}$ is the proportion of neighbor realizations with $i, j = 0$ (no exception) or 1 (exception). For example for two exceptions are neighbored, one has $i = j = 1$. Or if two realizations are not exceptions, then $i = j = 0$. These two ratios are both asymptotically $\chi^2(1)$ distributed.

The proposed ICVaR model is superior to the other two candidates. In the risk level test (LR1), the NIG fit performs even better than the HYP fit. In the HS-RM framework, the exceptions happen minimal 2.6 times and maximal 23 times more than the expected risk level, e.g. for the trading strategy $b = (1, 1)^\top$. In some cases, the underestimation is even over 25 times. Compared to the HS-RM, the HS-ES$t(df)$ method improves the VaR forecasting as the extreme risk levels such as 0.5% are considered. However both models are rejected in the two tests at 99% level.

An exemplary graphical illustration of the VaR forecasts is displayed in Figure 5.7, by which the resulting VaRs based on the ICVaR method and the other two alternative models are distinct in value. At this extremely risky situation, i.e. $\mathrm{pr} = 0.5\%$, the ICVaR based forecasts are accurate to inform the expected risk level with the empirical probability $\hat{\mathrm{pr}} = 6.5\%$. On the contrary, the HS-RM and HS-ES$t(df)$ models deliver many exceptions and underestimate the risk exposure over time. Their empirical risk levels are actually at $\hat{\mathrm{pr}} = 12.7\%$.

5.4.2 German stock portfolio

The raw data consists of 20 German stocks prices: Allianz, BASF, Bayer, BMW, Cobank, Daimler, Deutsche Bank, Degussa, Dresdner, Hoechst, Karstadt, Linde, Man, Mannesmann, Preussag, RWE, Schering, Siemens, Thyssen, Volkswagen, the largest and most liquid stocks traded on the Frankfurt German Exchange. The data spans 1974/01/02 to 1996/12/30, each has 5748 observations. It is available at FEDC as well. We choose a unit vector as the trading strategy for the whole period.

HYP	pr = 5%		pr = 1%		pr = 0.5%		NIG	pr = 5%		pr = 1%		pr = 0.5%	
day	mean	sd%	mean	sd%	mean	sd%	day	mean	sd%	mean	sd%	mean	sd%
1	-0.01	0.01	-0.02	0.04	-0.02	0.06	1	-0.01	0.02	-0.02	0.05	-0.02	0.06
2	-0.01	0.01	-0.01	0.04	-0.02	0.05	2	-0.01	0.01	-0.02	0.04	-0.02	0.06
.	.	.	.	.	.	.	.	.	.	.	.	.	.
424	-0.03	0.07	-0.06	0.15	-0.07	0.21	423	-0.04	0.06	-0.07	0.15	-0.07	0.23
425	-0.03	0.06	-0.06	0.14	-0.07	0.22	425	-0.04	0.06	-0.07	0.14	-0.07	0.24
426	-0.03	0.04	-0.05	0.11	-0.06	0.18	426	-0.03	0.05	-0.06	0.12	-0.06	0.18
597	-0.03	0.06	-0.06	0.16	-0.06	0.22	607	-0.04	0.06	-0.06	0.15	-0.07	0.21
.	.	.	.	.	.	.	.	.	.	.	.	.	.
999	-0.01	0.02	-0.02	0.04	-0.02	0.06	622	-0.03	0.07	-0.06	0.12	-0.07	0.18
1000	-0.01	0.02	-0.02	0.04	-0.02	0.07	1000	-0.01	0.02	-0.02	0.05	-0.02	0.07

Table 5.4: Descriptive statistics of daily empirical quantile estimates: $b = (1, 1)^\top$, MC simulation with $M = 10000$, $N = 100$, $T = 1000$.

HYP	pr = 5%		pr = 1%		pr = 0.5%		NIG	pr = 5%		pr = 1%		pr = 0.5%	
day	mean	sd%	mean	sd%	day	mean	sd%	mean	sd%	mean	sd%		
1	-0.01	0.02	-0.03	0.06	-0.03	0.09	1	-0.01	0.02	-0.03	0.06	-0.03	0.09
2	-0.01	0.02	-0.02	0.05	-0.03	0.08	2	-0.01	0.02	-0.03	0.06	-0.03	0.09
.	.	.	.	.	.	.	.	.	.	.	.	.	.
424	-0.05	0.10	-0.09	0.23	-0.10	0.31	325	-0.06	0.11	-0.10	0.22	-0.11	0.30
425	-0.05	0.10	-0.10	0.21	-0.10	0.33	423	-0.06	0.11	-0.10	0.25	-0.11	0.32
597	-0.05	0.09	-0.09	0.24	-0.10	0.33	424	-0.06	0.10	-0.10	0.22	-0.11	0.32
.	.	.	.	.	.	.	.	.	.	.	.	.	.
999	-0.01	0.03	-0.03	0.07	-0.03	0.08	597	-0.05	0.09	-0.10	0.22	-0.10	0.34
1000	-0.02	0.03	-0.03	0.06	-0.03	0.10	1000	-0.02	0.03	-0.03	0.08	-0.04	0.11

Table 5.5: Descriptive statistics of daily empirical quantile estimates: $b = (1, 2)^\top$, MC simulation with $M = 10000$, $N = 100$, $T = 1000$.

HYP	pr = 5%		pr = 1%		pr = 0.5%		NIG	pr = 5%		pr = 1%		pr = 0.5%	
day	mean	sd%	mean	sd%	mean	sd%	day	mean	sd%	mean	sd%	mean	sd%
1	-0.01	0.01	-0.01	0.03	-0.01	0.04	1	-0.01	0.01	-0.01	0.03	-0.02	0.05
2	-0.01	0.01	-0.01	0.03	-0.01	0.05	2	-0.01	0.01	-0.01	0.03	-0.02	0.05
.	.	.	.	.	.	.	.	.	.	.	.	.	.
601	-0.03	0.05	-0.06	0.15	-0.06	0.22	601	-0.03	0.05	-0.06	0.15	-0.07	0.19
607	-0.04	0.06	-0.06	0.13	-0.07	0.18	606	-0.03	0.06	-0.06	0.15	-0.07	0.23
611	-0.03	0.06	-0.06	0.13	-0.06	0.20	607	-0.04	0.07	-0.07	0.13	-0.07	0.19
613	-0.03	0.05	-0.06	0.15	-0.07	0.20	613	-0.04	0.06	-0.06	0.12	-0.07	0.18
.	.	.	.	.	.	.	.	.	.	.	.	.	.
999	-0.01	0.01	-0.01	0.03	-0.01	0.04	999	-0.01	0.01	-0.01	0.03	-0.01	0.04
1000	-0.01	0.01	-0.01	0.03	-0.01	0.04	1000	-0.01	0.01	-0.01	0.03	-0.01	0.05

Table 5.6: Descriptive statistics of daily empirical quantile estimates: $b = (-1, 2)^\top$, MC simulation with $M = 10000$, $N = 100$, $T = 1000$.

HYP	pr = 5%		pr = 1%		pr = 0.5%		NIG	pr = 5%		pr = 1%		pr = 0.5%	
day	mean	sd%	mean	sd%	mean	sd%	day	mean	sd%	mean	sd%	mean	sd%
1	-0.01	0.01	-0.01	0.03	-0.01	0.04	1	-0.01	0.01	-0.01	0.03	-0.02	0.04
2	-0.01	0.01	-0.01	0.03	-0.01	0.05	2	-0.01	0.01	-0.01	0.03	-0.02	0.05
.	.	.	.	.	.	.	.	.	.	.	.	.	.
607	-0.04	0.06	-0.06	0.13	-0.07	0.18	601	-0.03	0.06	-0.06	0.15	-0.07	0.22
608	-0.03	0.05	-0.06	0.13	-0.06	0.19	605	-0.04	0.06	-0.06	0.14	-0.07	0.19
612	-0.03	0.05	-0.06	0.14	-0.07	0.21	612	-0.04	0.07	-0.07	0.14	-0.07	0.19
.	.	.	.	.	.	.	.	.	.	.	.	.	.
999	-0.01	0.01	-0.01	0.03	-0.01	0.04	999	-0.01	0.01	-0.01	0.02	-0.01	0.04
1000	-0.01	0.01	-0.01	0.03	-0.01	0.04	1000	-0.01	0.01	-0.01	0.03	-0.02	0.04

Table 5.7: Descriptive statistics of daily empirical quantile estimates: $b = (-2, 1)^\top$, MC simulation with $M = 10000$, $N = 100$, $T = 1000$.

		ICVaR (HYP)			ICVaR (NIG)			HS-RM			HS-ES$t(df)$		
$b^\top$	pr%	p̂r%	LR1	LR2	p̂r%	LR1	LR2	p̂r%	LR1	LR2	p̂r%	LR1	LR2
(1,1)	5	6.7	5.52	*27.60	5.8	1.28	*28.82	10.8	*53.95	*19.84	11.0	*57.33	*26.05
(1,1)	1	1.9	6.47	*16.29	1.4	1.43	*11.14	5.5	*99.59	*21.70	5.3	*92.67	*22.37
(1,1)	0.5	0.9	2.59	5.58	0.7	0.71	0.00	5.0	*142.33	*23.44	4.2	*106.16	*18.96
(1,2)	5	6.8	6.16	*27.23	5.4	0.32	*30.50	10.9	*55.63	*24.78	11.0	*57.33	*24.48
(1,2)	1	1.6	3.07	*10.51	1.4	1.43	*11.14	5.3	*92.67	*22.37	4.9	*79.30	*23.81
(1,2)	0.5	0.9	2.59	0.00	0.8	1.52	0.00	4.7	*128.43	*21.77	4.4	*114.93	*22.83
(-1,2)	5	6.4	3.80	*38.56	5.5	0.51	*28.17	12.0	*75.40	*35.66	12.0	*75.40	*35.66
(-1,2)	1	1.4	1.43	*11.14	1.2	0.37	*11.86	6.0	*117.58	*25.86	5.5	*99.59	*23.40
(-1,2)	0.5	1.2	*7.06	*11.86	1.1	5.38	*15.84	4.4	*114.93	*18.33	3.2	*65.54	*20.90
(-2,1)	5	6.5	4.34	*22.35	5.0	0.00	*20.34	12.7	*89.18	*18.41	12.7	*89.18	*18.41
(-2,1)	1	1.6	3.07	0.00	1.1	0.09	0.00	5.6	*103.12	*14.79	4.7	*72.87	4.49
(-2,1)	0.5	0.9	2.59	0.00	0.8	1.52	0.00	4.2	*106.16	2.40	3.2	*65.54	5.54

Table 5.8: Backtesting of the VaR forecast of the exchange portfolios: MC simulation with $M = 10000$, $N = 100$, $T = 1000$. * indicates the model is rejected at 99% confidence level.

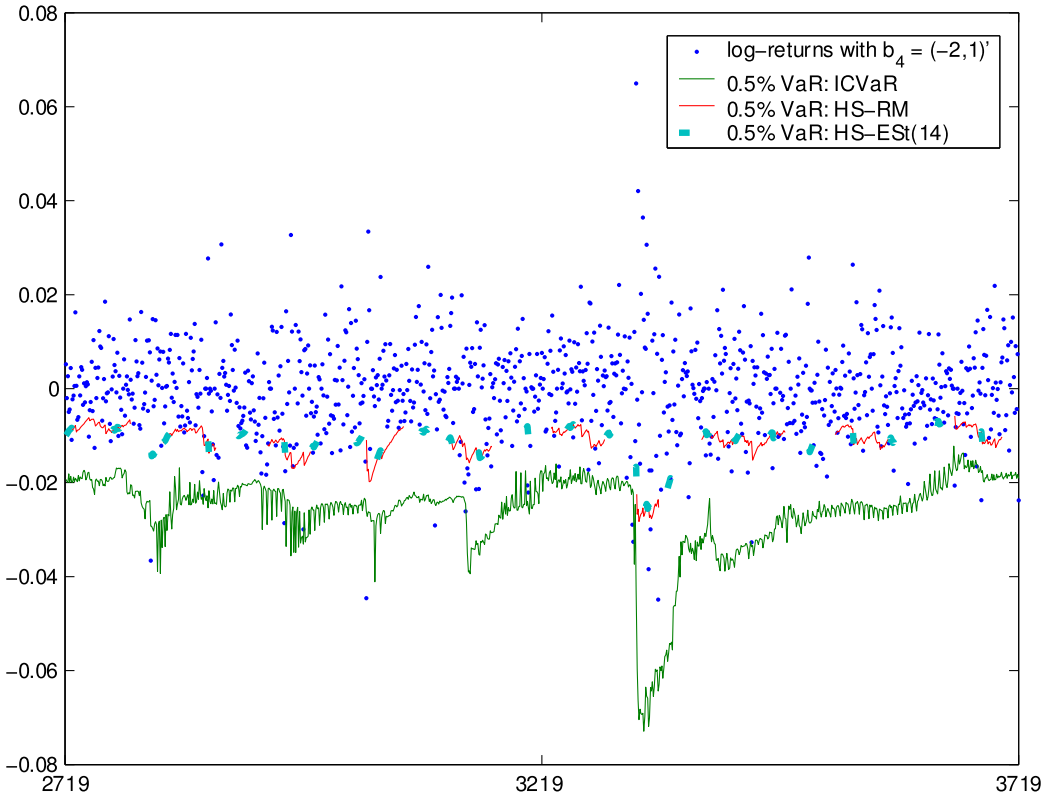

Figure 5.7: VaR time plot of the exchange rate portfolio with trading strategy $b_4 = (-2, 1)^\top$ at risk level pr $= 0.5\%$. Three risk management models are implemented: ICVaR (HYP), HS-RM and HS-ESt(14).

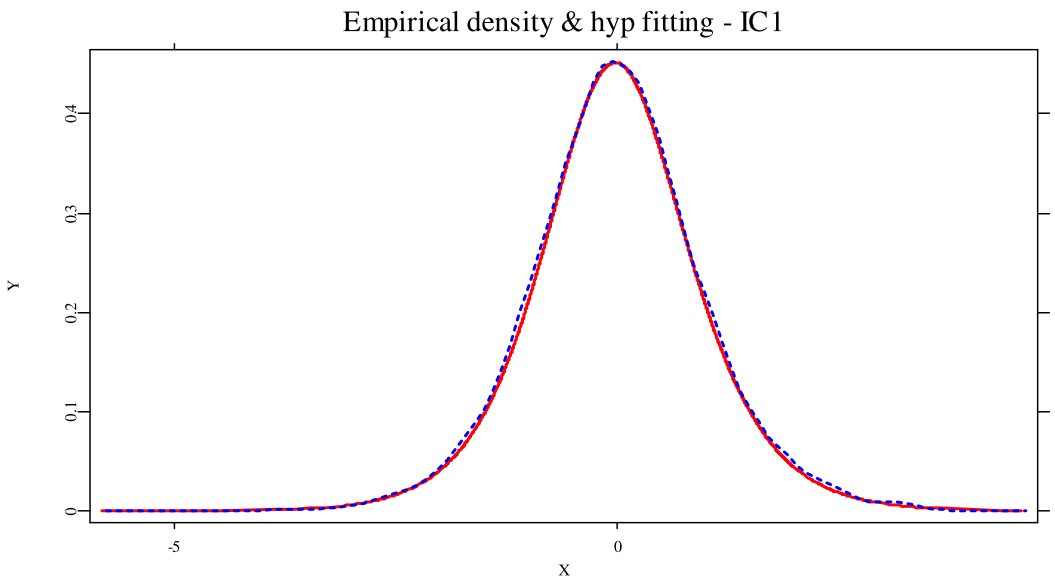

Figure 5.8: Density estimation of the first IC on the basis of the German stock portfolio. The HYP fit is displayed as a straight curve and the nonparametric density estimation is plotted by a dotted curve.

Table 5.9 summarizes the statistical properties of the log returns and the ICs. The log returns have sample means 0, their skewness are around 0 and the kurtoses are larger than 3. In this sense, the individual returns are approximately symmetric and leptokurtic. Furthermore, the auto-correlations of the first two orders ρ_1 and ρ_2 are small, indicating a negligible temporal dependency of every series. The smallest correlation is however 0.37, between the stocks Allianz (a global insurer and provider of financial services) and Preussag (a diversified German mining company and since 1997 known as TUI, a tourism company), indicating linear correlations among all the series. On the contrary, the correlations of the ICs are 0. The estimated ICs are leptokurtic and low temporal dependence of each IC series is observed. The GHADA method is applied to estimate the local volatility and identify the HYP or NIG distributional parameters for each IC. The distributional parameters are listed in Table 5.10. In order to check the quality of the parameters estimated, we compare the nonparametric density and the HYP fit of the estimated ICs. The exemplary graphical comparisons of the NIG fit are displayed in Figure 5.8 and Figure 5.9 respectively. It is clear that the GH identification well matches the empirical distributional behavior of the estimated IC. The resulting GH parameters are used in the MC simulation.

returns	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
variance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
skewness	0.07	-0.17	0.04	-0.01	-0.24	0.03	-0.30	-0.39	0.12	-0.12	-0.40	-0.24	-0.59	-0.27	0.13	-0.19	-0.03	-0.54	-0.05	-0.32
kurtosis	32.40	8.65	9.60	17.02	10.03	26.67	13.77	19.12	8.82	9.98	20.43	14.56	18.03	13.69	10.34	16.72	9.57	10.30	6.10	10.23
ρ_1	-0.06	-0.02	-0.02	-0.01	0.04	-0.02	0.03	0.02	0.06	0.01	0.00	-0.03	0.02	0.05	0.08	0.04	0.06	0.06	0.04	0.05
ρ_2	-0.02	-0.04	-0.04	-0.01	-0.04	-0.01	-0.03	-0.01	-0.03	-0.04	0.00	-0.01	-0.01	-0.01	-0.05	-0.02	-0.01	-0.01	-0.05	-0.01
ICs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
mean	0.01	-0.00	0.01	0.01	0.00	0.00	0.00	0.01	-0.00	0.01	0.01	-0.00	-0.01	-0.03	-0.00	-0.01	-0.00	-0.02	-0.00	0.00
variance	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
skewness	-1.44	-0.81	0.71	0.8	1.01	-0.57	0.68	0.38	-0.84	0.48	0.52	-0.04	-0.04	-0.24	-0.17	-0.51	-0.05	-0.13	0.06	-0.32
kurtosis	59.71	46.51	57.51	43.72	81.20	13.92	25.31	27.28	14.84	14.30	14.40	9.12	11.89	6.26	6.94	9.34	7.08	6.48	5.62	6.15
ρ_1	0.00	-0.26	-0.26	-0.11	-0.12	0.10	-0.02	-0.13	-0.01	0.06	-0.12	0.08	0.04	0.04	0.08	0.08	0.08	0.05	-0.01	0.07
ρ_2	-0.01	0.02	0.02	0.00	0.00	0.01	0.00	-0.01	-0.01	0.00	-0.01	0.01	0.02	-0.01	0.00	-0.03	0.02	0.00	0.00	0.00

Table 5.9: Descriptive statistics of the 20-dimensional German stock portfolio.

ICs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
HYP $\hat{\alpha}$	2.01	1.54	1.63	1.63	1.59	2.19	1.64	1.57	1.59	1.83	1.78	1.89	1.96	1.76	1.70	1.95	1.76	1.85	1.91	1.79
HYP $\hat{\beta}$	0.07	-0.04	0.07	0.11	0.13	0.09	0.14	0.02	-0.10	0.04	0.07	0.09	-0.00	-0.12	-0.11	-0.21	-0.03	-0.07	-0.00	-0.20
HYP $\hat{\delta}$	1.12	0.42	0.38	0.35	0.30	1.34	0.38	0.30	0.41	0.77	0.61	0.91	0.96	0.72	0.53	0.91	0.65	0.81	0.90	0.72
HYP $\hat{\mu}$	-0.05	0.02	-0.04	-0.10	-0.12	-0.08	-0.12	-0.01	0.10	-0.03	-0.06	-0.09	-0.01	0.09	0.11	0.17	0.02	0.04	-0.00	0.20
NIG $\hat{\alpha}$	1.63	0.97	0.98	0.95	0.90	1.83	0.95	0.90	1.00	1.34	1.21	1.45	1.54	1.29	1.15	1.55	1.25	1.35	1.44	1.34
NIG $\hat{\beta}$	0.05	-0.03	0.06	0.10	0.11	0.06	0.12	0.02	-0.09	0.04	0.07	0.06	0.00	-0.10	-0.10	-0.22	-0.02	-0.06	-0.00	-0.19
NIG $\hat{\delta}$	1.56	1.00	0.90	0.86	0.83	1.73	0.86	0.85	0.97	1.24	1.07	1.37	1.42	1.23	1.05	1.38	1.15	1.26	1.34	1.23
NIG $\hat{\mu}$	-0.03	0.02	-0.04	-0.09	-0.10	-0.04	-0.11	-0.01	0.09	-0.03	-0.06	-0.06	-0.02	0.07	0.10	0.18	0.01	0.02	-0.00	0.19

Table 5.10: HYP and NIG parameters estimates of the German stock portfolio.

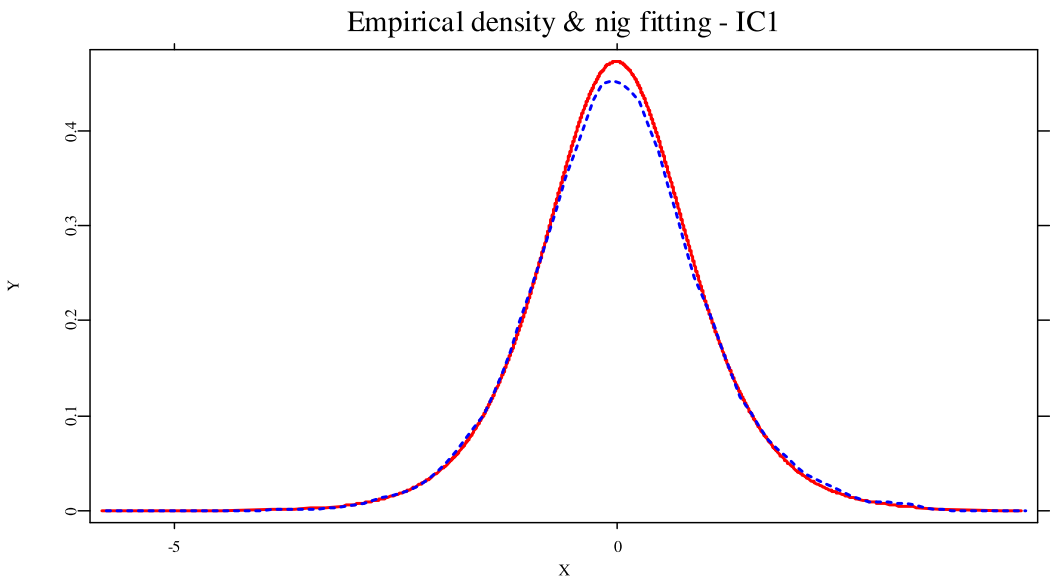

Figure 5.9: Density estimation of the first IC on the basis of the German stock portfolio. The NIG fit is displayed as a straight curve and the nonparametric density estimation is plotted by a dotted curve.

For the last $T = 1000$ days, we generate $d = 20$ samples with $M = 10,000$ observations. This scenario is repeated $N = 100$ times. The mean and the standard deviation of the daily empirical quantiles at the 5 risk levels are summarized in Table 5.11 and Table 5.12. The largest sd% w.r.t. the risk levels are underlined. The small values of deviation are evident to show the efficiency of the VaR inferences.

Figure 5.10 displays the one-day-ahead VaR forecasts for the German stock portfolio at pr = 5% and pr = 0.5%. The VaRs are calculated based on the ICVaR (NIG), HS-RM and the HS-ESt(19) methods. The df of the Student-t distribution 19 is estimated by the ML method. The exceptions of the proposed ICVaR methodology are obviously less than those given by the HS-RM and the HS-ES$t(df)$. A detailed comparison is based on the backtesting results in Table 5.13, where the HS-RM and the HS-ES$t(df)$ methods are rejected at 99% level in the risk level test. Although the ICVaR with HYP distributional assumption is rejected at extreme risk levels, i.e. $0.5\%, 0.25\%$ and 0.1% as well, it is still superior to the two methods mentioned above. On the meanwhile, the ICVaR with NIG distributional assumption is successful in all the risk level tests. Since the degree of freedom in the HS-ES$t(df)$ model has a large influence on the accuracy of forecast, we calculate the

HYP	pr = 5%		pr = 1%		pr = 0.5%		pr = 0.25%		pr = 0.1%	
day	mean	sd%	mean	sd%	mean	sd%	mean	sd%	mean	sd%
1	-0.28	0.36	-0.43	0.66	-0.45	0.92	-0.50	1.33	-0.56	1.91
2	-0.27	0.35	-0.41	0.76	-0.43	0.93	-0.48	1.25	-0.54	2.04
.	.	.	.	.	.	.	.	.	.	.
993	-0.38	0.48	-0.59	1.11	-0.63	1.41	-0.70	1.98	-0.80	3.34
.	.	.	.	.	.	.	.	.	.	.
997	-0.40	0.54	-0.64	1.08	-0.67	1.58	-0.75	2.15	-0.84	3.18
998	-0.38	0.55	-0.59	0.96	-0.63	1.56	-0.70	2.08	-0.79	3.22
999	-0.32	0.46	-0.49	0.89	-0.52	1.39	-0.58	1.90	-0.65	2.75
1000	-0.35	0.49	-0.54	0.94	-0.57	1.33	-0.63	1.75	-0.72	2.89

Table 5.11: Descriptive statistics of quantile estimates: MC simulation with $M = 10000$, $N = 100$, $T = 1000$.

NIG	pr = 5%		pr = 1%		pr = 0.5%		pr = 0.25%		pr = 0.1%	
day	mean	sd%	mean	sd%	mean	sd%	mean	sd%	mean	sd%
1	-0.30	0.42	-0.47	0.82	-0.49	1.13	-0.55	1.55	-0.62	2.15
2	-0.29	0.40	-0.44	0.79	-0.47	0.93	-0.52	1.35	-0.59	1.96
.	.	.	.	.	.	.	.	.	.	.
993	-0.40	0.58	-0.64	1.17	-0.68	1.55	-0.76	2.15	-0.87	3.75
.	.	.	.	.	.	.	.	.	.	.
997	-0.44	0.71	-0.69	1.55	-0.73	2.14	-0.81	2.61	-0.92	3.42
998	-0.41	0.59	-0.64	1.20	-0.68	1.58	-0.76	1.91	-0.86	3.49
999	-0.34	0.50	-0.53	0.96	-0.56	1.18	-0.62	1.71	-0.70	2.42
1000	-0.38	0.53	-0.59	0.99	-0.62	1.38	-0.69	1.90	-0.78	2.87

Table 5.12: Descriptive statistics of quantile estimates: MC simulation with $M = 10000$, $N = 100$, $T = 1000$.

VaRs by using different degrees of freedom ($df \in [4, 24]$) in the HS-ES$t(df)$ model. The HS-ES$t(df)$ on general performs worse than the ICVaR except one design with $df = 4$ at pr $= 0.1\%$, it does even better than the ICVaR (NIG). However, this model is not desirable in practice since it overestimates the risk and induce high capital reserve, see Figure 5.11 Moreover, the HS-ES$t(4)$ is rejected at the other four risk levels and hence it is not consistent to apply.

5.5 Conclusion

In this chapter, we proposed an easy and fast multivariate risk management model. The study is mainly based on the ICA. Instead of estimating the joint density and covariance of high-dimensional returns, the searching of ICs transfers the calculation to unidimensional studies. In the empirical study, the proposed ICVaR is superior to the RiskMetrics and t-deGARCH methods, above all in the risk level controlling. In addition, in the ICVaR methodology, the joint distribution of portfolio does not rely on trading strategy and therefore can be further applied to calculate VaRs as the investing positions change. Moreover, the ICVaR approach can be easily applied to calculate and forecast other risk measures such as expected shortfall.

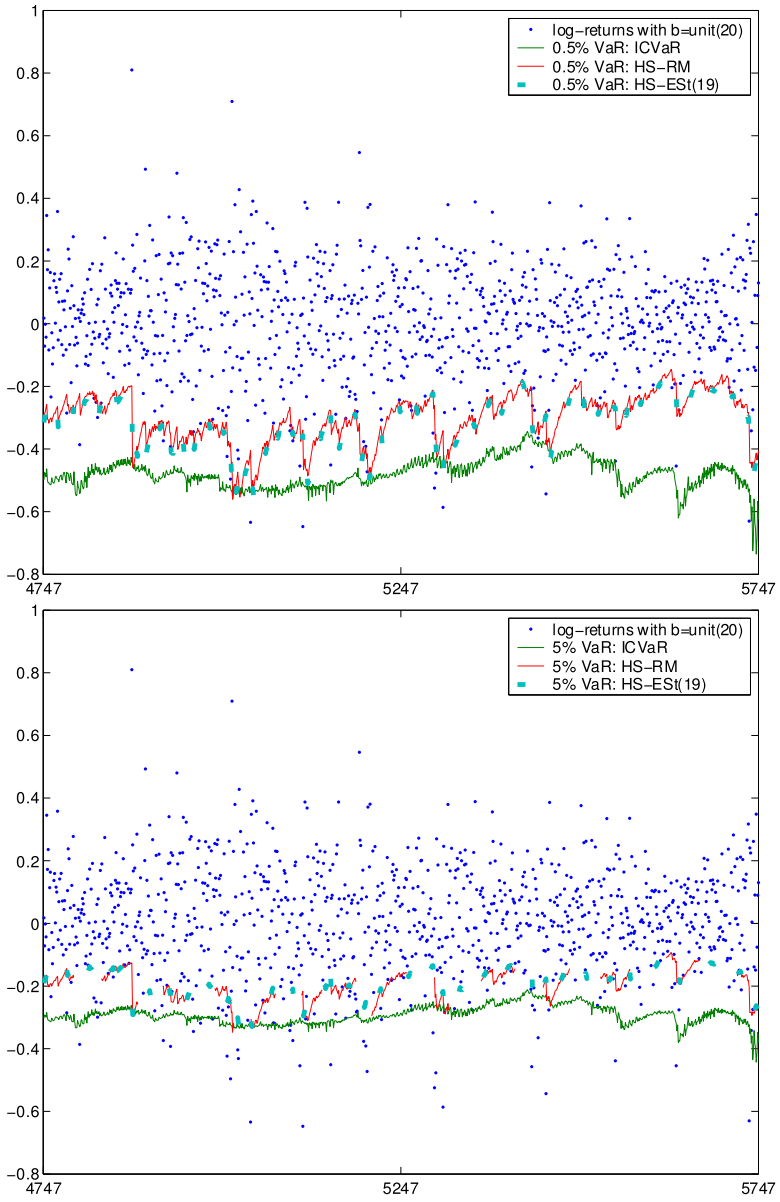

Figure 5.10: VaR time plots of the German stock portfolio with the equal weights. Three risk management models are implemented: ICVaR (NIG), HS-RM and HS-ESt(19). The risk levels are respectively pr $= 5\%$ (top) and 0.5% (bottom).

	pr = 5%			pr = 1%			pr = 0.5%			a = 0.25%			a = 0.1%		
	p̂r%	LR1	LR2	p̂r%	LR1	LR2	p̂r%	LR1	LR2	p̂r%	LR1	LR2	p̂r%	LR1	LR2
HYP	4.9	0.02	*16.76	1.7	4.09	0.13	1.2	*7.06	0.04	0.8	*7.64	0.03	0.7	*15.27	0.00
NIG	3.8	3.29	*7.14	1.4	1.43	0.05	0.8	1.52	0.03	0.7	5.43	0.00	0.4	5.09	0.00
HS-RM	10.4	*47.46	*18.29	5.4	*96.11	*7.28	3.5	*77.12	3.92	2.8	*84.94	0.42	2.0	*82.19	0.15
HS-ESt(4)	12.7	*89.18	*12.28	3.2	*30.93	4.27	1.7	*17.75	0.13	1.2	*18.73	0.04	0.2	0.77	0.00
HS-ESt(5)	12.2	*79.24	*14.25	3.3	*33.33	4.15	2.0	*25.67	0.15	1.3	*21.97	0.05	0.6	*11.52	0.00
HS-ESt(6)	11.7	*69.77	*16.40	3.5	*38.33	3.92	2.1	*28.53	0.16	1.3	*21.97	0.05	0.9	*23.61	0.00
HS-ESt(7)	11.4	*64.32	*13.82	3.7	*43.56	6.50	2.2	*31.48	0.17	1.4	*25.37	0.10	1.1	*32.85	0.04
HS-ESt(8)	11.2	*60.78	*15.24	4.1	*54.68	5.88	2.4	*37.65	0.18	1.6	*32.58	0.12	1.2	*37.75	0.04
HS-ESt(9)	11.0	*57.33	*15.71	4.2	*57.59	5.74	2.5	*40.87	0.19	1.7	*36.38	0.13	1.3	*42.83	0.05
HS-ESt(14)	10.6	*50.66	*16.70	4.5	*66.61	5.34	2.9	*54.53	0.44	2.0	*48.48	0.15	1.3	*42.83	0.05
HS-ESt(19)	10.6	*50.66	*16.70	4.9	*79.30	*8.04	2.9	*54.53	0.44	2.1	*52.73	0.16	1.5	*53.43	0.11
HS-ESt(24)	10.6	*50.66	*16.70	4.9	*79.30	*8.04	3.2	*65.54	4.27	2.2	*57.07	0.17	1.6	*58.94	0.12

Table 5.13: Backtesting of the VaR forecasts of the German stock portfolios: MC simulation with $M = 10000$, $N = 100$, $T = 1000$. * indicates the rejection at 99% level.

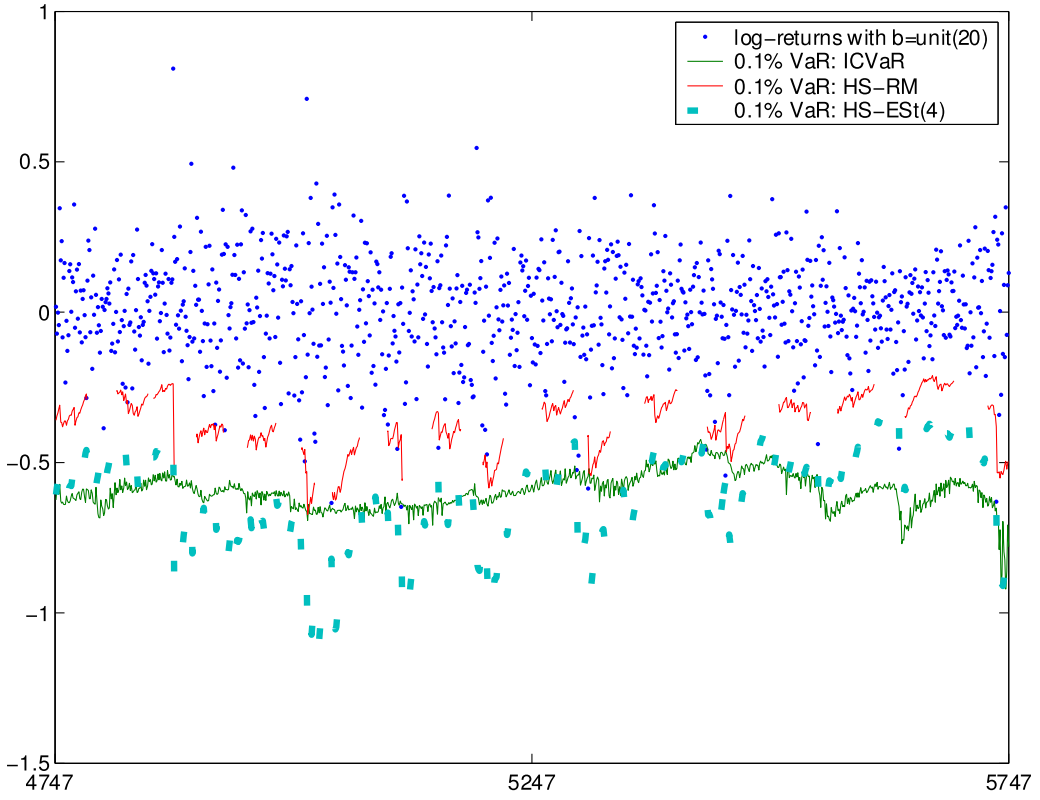

Figure 5.11: VaR time plots of the German stock portfolio with the equal weights. Three risk management models are implemented: ICVaR (NIG), HS-RM and HS-ESt(19). The risk level is pr = 0.1%.

Chapter 6

Adaptive risk management 4: GHICA

6.1 Introduction

For regulatory purposes, study on risk management is important for precise banking supervisory. Recent research provides detailed methodologies to calculate various risk measures. Among others, we refer to [Jor01] for a systematic description. Given a d-dimensional portfolio, the conditionally heteroscedastic model is widely used to describe the movement of the underlying series:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t), \tag{6.1}$$

where $x(t) \in \mathbb{R}^d$ are risk factors of the portfolio, e.g. (log) returns of the involved components. The covariance Σ_x is time dependent and needs to be assessed in a meaningful way. The stochastic innovations $\varepsilon_x(t) \in \mathbb{R}^d$ are assumed to be standardized with $\mathsf{E}[\varepsilon_x(t)|\mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\varepsilon_x^2(t)|\mathcal{F}_{t-1}] = I_d$. The popular risk measures are calculated based on the estimated joint density of the risk factors. For example, value at risk (VaR) is in fact the distributional quantile at a prescribed level over a target time interval and expected shortfall (ES) measures the size of losses once the realized losses exceed the VaR values. Indicated by formula (6.1), the joint density mainly relies on the covariance estimation and innovations' distributional assumption.

There are however some pitfalls and limitations when many risk management methods are applied to high dimensional portfolios. First risk measures are often calculated under the Gaussian distributional assumption, e.g. the RiskMetrics product introduced by JP Morgan in 1994. In the Gaussian framework with an estimate $\hat{\Sigma}_x(t)$ of $\Sigma_x(t)$, the standardized returns $\hat{\varepsilon}_x(t) = \hat{\Sigma}_x^{(-1/2)}(t)x(t)$ are asymptotically independent and the joint distri-

butional behavior can be therefore easily measured by the marginal distributions. However the Gaussian distributional assumption is merely used for computational and numerical purposes and not for statistical reasons. The conditional Gaussian marginal distributions and the resulting joint Gaussian distribution are at odds with empirical facts, i.e. financial series are heavy tailed distributed. The heavy tails are typically reduced but not eliminated as the series are standardized by the estimated volatility, see [ABDL01]. To alleviate the limitation, the Student-$t(6)$ distribution with degrees of freedom of 6 has been recommended. However this distribution often over-fits the heavy tails.

Figure 6.1 repeats the experiments in Figure 1.3 and demonstrates the effect of the distributional assumptions for two real data sets, the Allianz stock and a DAX portfolio from 1988/01/04 to 1996/12/30. The DAX is the leading index of Frankfurt stock exchange and a 20-dimensional hypothetic portfolio with a static trading strategy $b(t) = (1/20, \cdots, 1/20)^\top$ is considered. The portfolio returns $r(t) = b(t)^\top x(t)$ are analyzed in the univariate version of (6.1). This simplified calculation is used in practice, but it often suffers from low accuracy of calculation. Suppose now that the two return processes have been properly standardized, by using a local volatility estimation technique discussed later. The standardized returns are empirically heavy tailed distributed, indicated by the sample kurtoses 12.07 for the Allianz and 22.38 for the portfolio respectively. Three density estimations under the generalized hyperbolic (GH), Gaussian and $t(6)$ distributional assumptions are depicted in the figure. In the density comparison, the logarithmic density estimate using the nonparametric kernel estimation is considered as benchmark. The comparison w.r.t. the Allianz stock shows that the GH estimate is most close to the benchmark among others. The $t(6)$ estimate displays heavier tails relative to the benchmark, and the Gaussian estimate, on the contrary, presents lighter tails. The similar result is observed w.r.t. the DAX portfolio. It is therefore rational to surmise that the risk management methods under the Gaussian and $t(6)$ distributional assumptions generate low accurate results.

The second limitation of the popular multivariate risk management methods is due to high computational demand. Above all, these methods concern the covariance estimation to recover the empirical behavior of high dimensional series. Among many others, the constant conditional correlation (CCC) model proposed by [Bol90] and the subsequent dynamic conditional correlation (DCC) model proposed by [Eng02], [ES01] are very successful. In the estimation, the covariance matrix is approximated by the product of a diagonal matrix and a correlation matrix: $\Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top$, which reduces the number of unknown parameters relative to the BEKK specification proposed by [EK95b]. In spite of the appealing dimensional reduction,

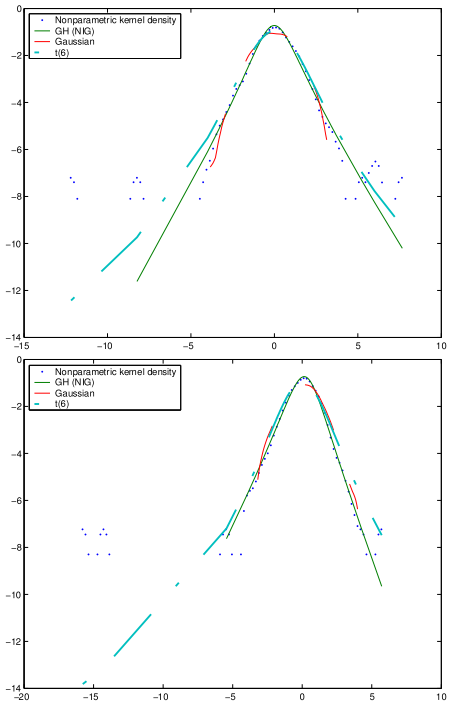

Figure 6.1: Density comparisons of the standardized returns in log scale based on the Allianz stock (top) and the DAX portfolio (bottom) with static weights $b(t) = \text{unit}(1/20)$. Time interval: 1988/01/04 - 1996/12/30. The nonparametric kernel density is considered as benchmark. The GH distributional parameters are respectively $\text{GH}(-0.5, 1.01, 0.05, 1.11, -0.03)$ for the Allianz and $\text{GH}(-0.5, 1.21, -0.21, 1.21, 0.24)$ for the DAX portfolio. Data source: FEDC (http://sfb649.wiwi.hu-berlin.de).

the mentioned estimation methods are still time consuming and numerically difficult to handle as really high dimensional series, e.g. a dimension $d > 10$, is considered, see [HHS03]. Moreover, these methods rely on the questionable Gaussian distributional assumption to ensure the independence of the resulting standardized returns. Otherwise, the distributional identification under a realistic assumption, such as the multivariate GH distribution with at least $4d$ parameters, involves once again numerical problem.

In accordance with the discussed limitations, [CHS6a] present a simple VaR calculation approach that achieves much better accuracy of calculation than the alternative RiskMetrics method. In their study, financial risk factors are first converted to independent components (ICs) using a linear filtering. The covariance matrix of the resulting ICs is a diagonal matrix and

the standardized ICs are independent and individually identifiable. By doing so, many univariate methods, that involve more realistic but complex procedures for local volatility estimation and distributional identification, can be easily applied to high dimensional series. [CHJ05], for example, propose a univariate VaR calculation approach by locally estimating volatility and well imitating the empirical distributional behavior of the underlying series. [CS06] present the local exponential smoothing method to estimate volatility and implement it in univariate risk management, by which the volatility estimates are sensitive to structure shifts and have low variability.

The primary aim of this paper is to introduce a simple and fast multivariate risk management method, **GHICA**, by implementing the IC analysis (ICA) to the high dimensional series and fitting the resulting ICs in the GH distributional framework. The GHICA method improves the work of [CHS6a] from two aspects. The volatility estimation is driven by the local exponential smoothing technique to achieve the best possible accuracy of estimation. The fast Fourier transformation (FFT) technique is used to approximate the density of the portfolio returns. Compared to the Monte Carlo simulation technique used in the former study, it significantly speeds up the calculation.

The proposed GHICA method is also applicable for covariance estimation. Relative to the widely used DCC setup, the GHICA method is fast and delivers sensitive estimates. This covariance estimation comparison is illustrated based on simulated data sets. Furthermore, the GHICA method is implemented to risk management on the base of DAX stocks and foreign exchange rates. Several hypothetic portfolios are constructed by assigning static and dynamic trading strategies to the data sets. The results are compared with the risk measures calculated by alternative methods, i.e. the RiskMetrics method, the method using the exponential smoothing to estimate volatility and assuming the $t(6)$ distribution, and the method using the DCC to estimate covariance in the Gaussian distributional framework. All the results are analyzed from the viewpoints of regulatory, investors and internal supervisory. The GHICA method, in general, produces better results than the others.

The paper is organized as follows. The GHICA method is described in Section 6.2, by which the ICA method, the local exponential smoothing technique and the FFT technique are detailed. Section 6.3 compares the covariance estimations using the GHICA and DCC methods. The simulated data are GH distributed with $d = 50$ components. The real data analysis in Section 6.4 demonstrates the quality of the GHICA method based on the 20-dimensional German DAX portfolios and a dynamic exchange rate portfolio. Several alternative methods are implemented as well to compare the accuracy of calculation with the GHICA one.

6.2 GHICA Methodology

Given multidimensional time series, for example prices of financial assets, $s(t) \in \mathbb{R}^d$, the (log) returns are calculated as $x(t) = \log\{s(t)/s(t-1)\}$. Without loss of generality, the drift of the returns is set to be 0. A time homogeneous model means that the covariance matrix $\Sigma_x(t)$ is a constant, i.e. $x(t) = \Sigma_x^{1/2}\varepsilon_x(t)$, where the innovations $\varepsilon_x(t) \in \mathbb{R}^d$ are assumed to be standardized with $\mathsf{E}[\varepsilon_x(t)|\mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\varepsilon_x^2(t)|\mathcal{F}_{t-1}] = I_d$. The maximum Gaussian likelihood estimate of the time homogeneous covariance Σ_x is the average value of the squared returns. Since the covariance is time dependent in practice, many techniques have been used to approximate the local maximum likelihood estimate (MLE) of the covariance in the conditional heteroscedastic model:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t),$$

by specifying a "local homogeneous" interval (e.g. one year or 250 trading days). Inside the homogeneous interval, the unknown covariance is almost time-invariant and can be identified using the ML estimation. Among many others, the multivariate GARCH setup such as the DCC is successful in characterizing the clustering feature of covariance under the Gaussian distributional assumption. Recall that the Gaussian assumption helps to ensure independence of the standardized innovations, which is essential for the ML estimation and the distributional identification. As the dimension d increases, it however has to estimate many parameters and becomes numerically difficult. Moreover, the standardized returns $\hat{\varepsilon}_x(t) = \hat{\Sigma}_x^{-1/2}(t)x(t)$ are empirically not Gaussian distributed. Under a realistic distributional assumption, on the other hand, the distributional behaviors such as asymmetry and heavy tails are well matched, but it is hard to identify the unknown distributional parameters due to complex density form.

The GHICA method proposes a solution to balance the numerical tractability and the realistic distributional assumption on the risk factors. It first converts the return series using the simple linear transformation and filters out ICs: $y(t) = Wx(t)$. The transformation matrix W is assumed to be time constant and nonsingular. The heteroscedastic model is reformulated as:

$$x(t) = W^{-1}y(t) = W^{-1}\Sigma_y^{1/2}(t)\varepsilon_y(t) = W^{-1}D_y^{1/2}(t)\varepsilon_y(t).$$

Due to the statistical property of independence, the covariance of the ICs $\Sigma_y(t)$ is a diagonal matrix and is denoted as $D_y(t)$ to emphasize this feature. Its diagonal elements are the time varying variances of the ICs. The stochastic innovations $\varepsilon_y(t) = \{\varepsilon_{y_1}(t), \cdots, \varepsilon_{y_d}(t)\}^\top$ are cross independent and can be individually identified in the realistic and univariate distributional frame-

work. By doing so, the GHICA method simplifies high dimensional analysis to univariate study and significantly speeds up the calculation.

In this section, the building blocks of the GHICA method are described. To be more specific, the FastICA procedure is used to estimate the transformation matrix W. The resulting ICs are individually analyzed, by which the univariate volatility process is estimated using the local exponential smoothing technique and the innovations are assumed to be GH distributed. The quantile of the portfolio return is calculated based on the FFT technique.

The GHICA algorithm is as follows:

1. Do ICA to the given risk factors to get ICs.
2. Implement local exponential smoothing to estimate the variance of each IC
3. Identify the distribution of every IC's innovation in the GH distributional framework
4. Estimate the density of the portfolio return using the FFT technique
5. Calculate risk measures

The second usage of the GHICA method is to estimate the covariance matrix $\Sigma_x(t)$ based on the matrix estimate $\hat{W}$ and the variance estimates of the ICs: $\hat{\Sigma}_x(t) = \hat{W}^{-1}\hat{D}_y(t)\hat{W}^{-1\top}$ An alternative covariance estimation approach, the DCC, is briefly described as well. We will compare the GHICA-based covariance estimation with the DCC estimation in the later simulation study.

6.2.1 Independent component analysis (ICA) and FastICA approach

The main aim of ICA is to retrieve, out of high dimensional time series, stochastically ICs through a linear transformation: $y(t) = Wx(t)$, where the transformation matrix $W = (w_1, \cdots, w_d)^\top$ is nonsingular. High order moments are essential for estimation of the ICs. In the Gaussian framework, higher order moments are however fixed such as skewness with value of 0 and kurtosis with value of 3. Therefore the ICs are assumed to be non-gaussian distributed. Furthermore, the above transformation holds true by simultaneously multiplying the same constants to the unknown terms $y(t)$ and W: $\{cy(t)\} = \{cW\}x(t)$. To avoid the scale identification problem, it is rational to standardize the dependent series and assume that every IC has unit variance $\mathsf{E}(y_j) = 1$ with $j = 1, \cdots, d$. The Mahalanobis transformation

$\tilde{x}(t) = \tilde{\Sigma}_x^{-1/2} x(t)$ helps to standardize the return series and the results are analyzed using the ICA:

$$y(t) = \tilde{W}\tilde{x}(t),$$

where $\tilde{\Sigma}_x$ is the sample covariance based on the available data. It is easy to show that the transformation matrix $\tilde{W}$ turns to be an orthogonal matrix with unit norm. The corresponding matrix w.r.t. the return series is $W = \tilde{W}\tilde{\Sigma}_x^{-1/2}$. For notational simplification, we eliminate the mark $\tilde{\cdot}$ in the following text in this section.

Various ideas have been proposed to estimate the transformation matrix W. Among them, one intuitive ICA estimation is motivated by the definition of mutual information. The mutual information is nonnegative and a natural measure of independence. It is defined as the difference of the sum of marginal entropy and the mutual entropy:

$$\begin{aligned} I(y) &= \sum_{j=1}^{d} H(y_j) - H(y) \qquad (6.2) \\ \text{where } H(y_j) &= -\int f_{y_j}(u) \log f_{y_j}(u) du \end{aligned}$$

Based on the linear transformation of the ICA, the mutual information in (6.2) can be reformulated as:

$$I(W, y) = \sum_{j=1}^{d} H(y_j) - H(x) - \log|\det(W)|.$$

Notice that the entropy of the return series $H(x)$ is a fixed value and does not depend on the ICs, and further the last term is 0 due to the orthogonality of the transformation matrix W. It is clear that the mutual information $I(\cdot)$ in (6.2) is equal to 0 if y_i and y_j are cross independent for $i \neq j$ and $i, j = 1, \ldots, d$, see [CT91]. Hence for a candidate transformation W, one can minimize the mutual information to achieve independence. The optimization problem is: $\min_W \sum_{j=1}^{d} H(y_j)$ and can be further simplified to d simple optimization problems according to the inequality:

$$\min_W \sum_{j=1}^{d} H(y_j) \;\geq\; \sum_{j=1}^{d} \min_{w_j} H(y_j)$$

This simplification leads to some loss in the W estimation but it extensively speeds up the estimation procedure by merely considering d elements of W every time. Moreover, the entropy and negentropy $J(y_j) = H(y_0) - H(y_j)$ are in one-to-one correspondence, where $y_0 \sim \mathrm{N}(0, 1)$ is a standard Gaussian

vector and $H(y_0)$ is merely a constant. The negentropy is always nonnegative since the Gaussian random variable has the largest entropy given the same variance, see [Hyv98]. Therefore, we can also formulate the optimization problem as:

$$\hat{w}_j = \text{argmin}H(y_j) = \text{argmax}J(w_j, y_j).$$

In the ICA estimation, the approximation of negentropy is used to construct the optimization object function w.r.t. the j-th row of the transformation matrix W:

$$\begin{aligned} \hat{w}_j &= \text{argmin}H(y_j) = \text{argmax}J(y_j) \\ J(y_j) &\approx const.\{\mathsf{E}[G(y)] - \mathsf{E}[G(y_0)]\}^2 \\ &= const.\{\mathsf{E}[G(w_j^\top x)] - \mathsf{E}[G(y_0)]\}^2 \\ G(y_j) &= \text{logcosh}(y_j) \end{aligned} \tag{6.3}$$

This optimization problem is solved by using the symmetric FastICA algorithm, see [HKO01]:

1. Initialization: Choose initial vectors $\hat{w}_j^{(1)}$ for $W = \{w_1, \cdots, w_d\}^\top$ with $j = 1, \cdots, d$, each has a unit norm.

2. Loop:

   - At step n, Calculate $\hat{w}_j^{(n)} = \mathsf{E}\left[x^\top(t) g\left\{\hat{w}_j^{(n-1)\top} x(t)\right\}\right] - \mathsf{E}\left[g'\left\{\hat{w}_j^{(n-1)\top} x(t)\right\}\right] \hat{w}_j^{(n-1)}$, where g is the first derivative of $G(y)$ in form (6.3) and g' is the second derivative. The expectation $\mathsf{E}[\cdot]$ is approximated by the sample mean.
   - Do a symmetric orthogonalization of the estimated transformation matrix $\hat{W}^{(n)}$:
   $$\hat{W}^{(n)} = \{\hat{W}^{(n)}\hat{W}^{(n)\top}\}^{-1/2}\hat{W}^{(n)}$$
   - If not converged, i.e. $\det\{\hat{W}^{(n)} - \hat{W}^{(n-1)}\} \neq 0$, go back to 2. Otherwise, the algorithm stops.

3. Final result: the last (converged) estimate is the final estimate $\hat{W}$.

6.2.2 Local exponential smoothing and dynamically conditional correlation

Suppose that the ICs and the transformation matrix W are given. The covariance matrices of the ICs and the original return series are respectively:

$$\begin{aligned} D_y(t) &= \text{diag}\{\sigma_{y_1}^2(t), \cdots, \sigma_{y_d}^2(t)\} \\ \Sigma_x(t) &= W^{-1} D_y(t) W^{-1\top} \end{aligned} \tag{6.4}$$

where $\sigma_{y_j}(t)$ is the heteroscedastic volatility of the j-th IC with $j = 1, \cdots, d$. Recall that (6.4) has a similar decomposition structure as the often used principal component analysis (PCA), by which the covariance is decomposed as: $\Sigma_x = \Gamma\Lambda\Gamma^\top$ with the eigenvector matrix Γ and the diagonal eigenvalue matrix Λ, see [Flu98]. Among other distinctions, the PCA method orders the resulting PCs whereas the ICs have equal importance. In the estimation of the unknown variance, the local exponential smoothing method is used to achieve the best possible accuracy of the volatility estimation.

Local exponential smoothing: Given the univariate conditional heteroscedastic model: $y_j(t) = \sigma_{y_j}(t)\varepsilon_{y_j}(t)$ with $\mathsf{E}[\varepsilon_{y_j}(t)|\mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\varepsilon^2_{y_j}(t)|\mathcal{F}_{t-1}] = 1$, we now focus on the adaptive estimation of the volatility σ_{y_j} for $j = 1, \cdots, d$. For notational simplification, the subscripts y_j in σ_{y_j} and j in y_j are eliminated here. Suppose that a finite set $\{\eta_k, k = 1, \cdots, K\}$ of values of smoothing parameter is given. Every value η_k leads to a localizing weighting scheme $\{\eta_k^{t-s}\}$ for $s \leq t$ to the local Gaussian MLE $\tilde{\sigma}^{(k)}(t)$ (In practice, one truncates the smoothing window at M_k such that $\eta_k^{M_k+1} \leq c \to 0$):

$$\tilde{\sigma}^{(k)}(t) \;=\; \left[\{\sum_{m=0}^{M_k} \eta_k^m y^2(t-m-1)\}/\{\sum_{m=0}^{M_k} \eta_k^m\}\right]^{1/2}$$

As discussed before, financial time series are heavy tailed distributed and can be well identified in the GH distributional framework. Under the normal inverse Gaussian (NIG) distributional assumption, (one subclass of the GH distribution, see Section 6.2.3 for more details), the quasi ML estimation is applicable if the exponential moment of the squared innovations $\mathsf{E}[\exp\{\rho\varepsilon^2(t)\}]$ exists. A power transformation guarantees that:

$$\begin{aligned} y_p(t) &= \text{sign}\{y(t)\}|y(t)|^p \\ \theta(t) &= Var\{y_p(t)|\mathcal{F}_{t-1}\} = \mathsf{E}\{y_p^2(t)|\mathcal{F}_{t-1}\} = \mathsf{E}\{|y(t)|^{2p}|\mathcal{F}_{t-1}\} \\ &= \sigma^{2p}(t)\,\mathsf{E}\,|\varepsilon(t)|^{2p} = \sigma^{2p}(t)C_p \end{aligned} \tag{6.5}$$

where $C_p = \mathsf{E}(|\varepsilon(t)|^{2p}|\mathcal{F}_{t-1})$ is a constant and only relies on $0 \leq p < 1/2$. Notice that the power transformed variable $\theta(t)$ is one-to-one correspondence to the variance $\sigma^2(t)$ and can be estimated on the base of the transformed observations $|y(t)|^{2p}$:

$$\begin{aligned} \tilde{\theta}^{(k)}(t) &= \{\sum_{m=0}^{M_k} \eta_k^m |y(t-m-1)|^{2p}\}/\{\sum_{m=0}^{M_k} \eta_k^m\} \\ \text{with } N_k &= \sum_{m=0}^{M_k} \eta_k^m \end{aligned} \tag{6.6}$$

The smoothing parameter η_k runs over a wide range from values close to zero to one, so that the variability of the unknown process $\theta(t)$ reduces and at least one of the resulting MLEs is good in the sense of small estimation bias. [PS06] show that the inverse of N_k in (6.6) is positively related to the variation of the MLEs. This result is used to construct the sequence of the smoothing parameter $\{\eta_k\}$:

$$\frac{N_{k+1}}{N_k} \approx \frac{1-\eta_k}{1-\eta_{k+1}} = a > 1, \tag{6.7}$$

where the coefficient a controls the decreasing speed of the variations.

The corresponding fitted log-likelihood ratio $L\left(\eta_k, \tilde{\theta}^{(k)}(t), \theta(t)\right)$ reads as:

$$L\left(\eta_k, \tilde{\theta}^{(k)}(t), \theta(t)\right) = L(\eta_k, \tilde{\theta}^{(k)}(t)) - L(\eta_k, \theta(t)) = N_k \mathcal{K}\left(\tilde{\theta}^{(k)}(t), \theta(t)\right)$$

where the nonnegative Kullback-Leibler divergence is:

$$\mathcal{K}\left(\tilde{\theta}^{(k)}(t), \theta(t)\right) = -\frac{1}{2}\{\log\frac{\tilde{\theta}^{(k)}(t)}{\theta(t)} - \frac{\tilde{\theta}^{(k)}(t)}{\theta(t)} + 1\}.$$

The local MLEs $\tilde{\theta}^{(k)}(t)$ will be referred as "weak" estimates since the final estimate $\hat{\theta}(t)$ achieves the best possible accuracy of estimation by aggregating all the weak estimates.

The procedure is sequential and starts with the estimate $\tilde{\theta}^{(1)}(t)$ that has the largest variability but small bias, i.e. we set $\hat{\theta}^{(1)}(t) = \tilde{\theta}^{(1)}(t)$. At every step $k \geq 2$, the new estimate $\hat{\theta}^{(k)}(t)$ is constructed by aggregating the next "weak" estimate $\tilde{\theta}^{(k)}(t)$ and the previously constructed estimate $\hat{\theta}^{(k-1)}(t)$. Following to [BS06], the aggregation is done in terms of the parameter $v = -1/(2\theta)$ so that the variable $y(t)$ belongs to the exponential distributional family with a density form: $p(y, v) = p(y)\exp\{yv - d(v)\}$:

$$\begin{aligned} \hat{v}^{(k)}(t) &= \gamma_k \tilde{v}^{(k)}(t) + (1-\gamma_k)\hat{v}^{(k-1)}(t) \\ \text{or equivalently, } \hat{\theta}^{(k)}(t) &= \left(\frac{\gamma_k}{\tilde{\theta}^{(k)}(t)} + \frac{1-\gamma_k}{\hat{\theta}^{(k-1)}(t)}\right)^{-1} \end{aligned}$$

The mixing weights $\{\gamma_k\}$ are computed on the base of the fitted log-likelihood ratio by checking that the previously aggregated estimate $\hat{\theta}^{(k-1)}(t)$ is in agreement with the next "weak" estimate $\tilde{\theta}^{(k)}(t)$, i.e. the difference between these two estimates is bounded by critical values $\mathfrak{z}_k$:

$$\begin{aligned} \gamma_k &= K_{ag}\left\{2L\left(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t)\right)/\mathfrak{z}_k\right\} \\ \text{where } K_{ag}(u) &= \{1 - (u - 1/6)_+\}_+ \end{aligned}$$

The aggregation kernel K_{ag} guarantees that the mixing coefficient γ_k is one if there is no essential difference between $\tilde{\theta}^{(k)}(t)$ and $\hat{\theta}^{(k-1)}(t)$, and zero if the difference is significant. The significance level is measured by the critical value ζ_k. In [CS06], the critical values for $p = 0.25$ have been calculated:

$$\mathfrak{z}_k = 0.008 + 0.005 * (K - k).$$

In the intermediate case, the mixing coefficient γ_k is between zero and one. The procedure terminates after step k if $\gamma_k = 0$ and we define in this case $\hat{\theta}^{(m)}(t) = \hat{\theta}^{(k-1)}(t)$ for all $m \geq k$.

The corresponding volatility estimates $\hat{\sigma}(t)$ are according to (6.5):

$$\hat{\sigma}(t) \quad = \quad \{\hat{\theta}(t)/C_p\}^{\frac{1}{2p}},$$

where the constant C_p is computed such that the residuals $\hat{\varepsilon}(t) = y(t)/\hat{\sigma}(t)$ have a unit variance as assumed in the heteroscedastic model.

The algorithm is described as follows:

1. Initialization: $\hat{\theta}^{(1)}(t) = \tilde{\theta}^{(1)}(t)$.
2. Loop: for $k \geq 2$

$$\hat{\theta}^{(k)}(t) = \left(\frac{\gamma_k}{\tilde{\theta}^{(k)}(t)} + \frac{1 - \gamma_k}{\hat{\theta}^{(k-1)}(t)} \right)^{-1}$$

   where the mixing coefficient γ_k is computed as:

$$\gamma_k = K_{ag} \left\{ 2L \left(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t) \right) / \mathfrak{z}_k \right\}$$

3. Final estimate: if $k = K$, $\hat{\theta}(t) = \hat{\theta}^{(K)}(t)$ or $\gamma_k = 0$, $\hat{\theta}(t) = \hat{\theta}^{(k-1)}(t)$.
4. Compute C_p such that $Var\{\hat{\varepsilon}(t)\} = Var\left[y(t) \{C_p/\hat{\theta}(t)\}^{\frac{1}{2p}} \right] = 1$.
5. Compute volatility estimate $\hat{\sigma}(t) = \{\hat{\theta}(t)/\hat{C}_p\}^{\frac{1}{2p}}$.

Consequently, the covariance matrices $D_y(t)$ and $\Sigma_x(t)$ are calculated.

Dynamic conditional correlation (DCC) model: Alternatively, the covariance of the return series can be estimated by the DCC model:

$$\Sigma_x(t) = D_x(t) R_x(t) D_x(t)^\top.$$

This technique first identifies the elements of the diagonal matrix $D_x(t)$ in the GARCH(1,1) setup and adaptively specifies the correlation matrix as:

$$R_x(t) = \tilde{R}_x(1 - \theta_1 - \theta_2) + \theta_1 \{\varepsilon_x(t-1)\varepsilon_x(t-1)^\top\} + \theta_2 R_x(t-1),$$

where $\tilde{R}_x$ is the sample correlation of the risk factors, $\varepsilon_x \in \mathbb{R}^d$ are the standardized returns, i.e. risk factors divided by the univariate GARCH(1,1) volatilities, or equivalently by the squared diagonal elements in $D_x(t)$. The standardized returns are assumed to be Gaussian distributed. The parameters θ_1 and θ_2 are identified by the ML estimation.

6.2.3 Normal inverse Gaussian (NIG) distribution and fast Fourier transformation (FFT)

After the ICA, one obtains ICs that are assumed to be NIG distributed. The NIG is a subclass of the GH distribution with a fixed value of $\lambda = -1/2$, see [EP02]. With 4 distributional parameters, the NIG distribution is flexible to well match the behavior of real data. Compared to many other subclasses of GH distribution, the NIG distribution has a desirable property, saying that the scaled NIG variable belongs to the NIG distribution as well. The density of NIG random variable has a form of:

$$f_{\text{NIG}}(y;\alpha,\beta,\delta,\mu) = \frac{\alpha\delta}{\pi}\frac{K_1\left\{\alpha\sqrt{\delta^2+(y-\mu)^2}\right\}}{\sqrt{\delta^2+(y-\mu)^2}}\exp\{\delta\sqrt{\alpha^2-\beta^2}+\beta(y-\mu)\},$$

where the distributional parameters fulfill $\mu \in \mathbb{R}$, $\delta > 0$ and $|\beta| \leq \alpha$. The modified Bessel function of the third kind $K_\lambda(\cdot)$ with an index $\lambda = 1$ has a form of:

$$K_\lambda(y) = \frac{1}{2}\int_0^\infty y^{\lambda-1}\exp\{-\frac{y}{2}(y+y^{-1})\}\,dy$$

The characteristic function of the NIG variable is:

$$\varphi_y(z) = \exp\left[\mathbf{i}z\mu + \delta\{\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+\mathbf{i}z)^2}\}\right]$$

Proof: The characteristic function of the GH random variable has a form of:

$$\varphi_y(z) = \exp(\mathbf{i}z\mu)\left\{\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+\mathbf{i}z)^2}\right\}^{\lambda/2}\frac{K_\lambda\{\delta\sqrt{\alpha^2-(\beta+\mathbf{i}z)^2}\}}{K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}$$

Using the representation of the modified Bessel function with a fixed index $\lambda = -1/2$ derived in [BNB81]:

$$K_\lambda(y) = \sqrt{\frac{2}{\pi}}y^{-1/2}e^{-y},$$

it is straightforwardly to show that the assertion holds. □

One desirable feature of the NIG distribution is its explicit scaling transformation. Multiplying the random variable by c, the resulting variable $y' = cy$ belongs to the NIG distribution as well:

$$f_{\text{NIG}}(y'; \alpha', \beta', \delta', \mu') = f_{\text{NIG}}(cy; \alpha/|c|, \beta/c, |c|\delta, c\mu). \tag{6.8}$$

Proof: It is easy to show the result by using the Jacobian transformation, see [HS03]. Given the density of y and let $\alpha' = \alpha/|c|$, $\beta' = \beta/c$, $\delta' = |c|\delta$ and $\mu' = c\mu$, the density of $y' = cy$ has a form of:

$$\begin{aligned} f(y') &= \frac{1}{|c|} f_y(\frac{y}{c}) = \frac{\alpha'\delta'}{\pi} \frac{K_1\left\{\alpha'\sqrt{\delta'^2 + (y'-\mu')^2}\right\}}{\sqrt{\delta'^2 + (y'-\mu')^2}} \\ &\quad \exp\{\delta'\sqrt{\alpha'^2 - \beta'^2} + \beta'(y'-\mu')\} = f_{\text{NIG}}(y'; \alpha', \beta', \delta', \mu'). \end{aligned}$$

□

To calculate risk measures, it requires the identification of the portfolio returns' density. Based on the GHICA model, the portfolio returns are formulated as:

$$r(t) = b(t)^\top W^{-1} D_y(t)^{1/2} \varepsilon_y(t)$$

where $b(t)$ is the trading strategy. Notice that the linear transformation of the NIG variable is not necessarily NIG distributed. In other words, the density of the return is unknown although the marginal densities are clear. On the meanwhile its characteristic function is explicitly writable. This is the same case as approximating the α-stable distribution in [MR04], by which the Fourier transformation is used to approximate the density of the variable based on its characteristic function. This motivates us to use the technique to approximate the density of the return in the GHICA procedure.

Set $a = (a_1, \cdots, a_d) = b(t)^\top W^{-1} D_y(t)^{1/2}$, the variable $\zeta_j = a_j \varepsilon_j$ is NIG distributed with $j = 1, \cdots, d$, according to (6.8):

$$\zeta_j \sim \text{NIG}(\zeta_j, \breve{\alpha}_j, \breve{\beta}_j, \breve{\delta}_j, \breve{\mu}_j) = \text{NIG}(\zeta_j, \alpha_j/|a_j|, \beta_j/a_j, |a_j|\delta_j, a_j\mu_j).$$

The characteristic function of the return $r = \sum_{j=1}^d \zeta_j$ at time t is:

$$\varphi_r(z) = \prod_{j=1}^d \varphi_{\zeta_j}(z) = \exp\left[\mathbf{i}z \sum_{j=1}^d \breve{\mu}_j + \sum_{j=1}^d \breve{\delta}_j \{\sqrt{\breve{\alpha}_j^2 - \breve{\beta}_j^2} - \sqrt{\breve{\alpha}_j^2 - (\breve{\beta}_j + \mathbf{i}z)^2}\}\right]$$

The density function is approximated by the Fourier transformation:

$$f(r) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-\mathbf{i}tr)\psi(z)dt \approx \frac{1}{2\pi} \int_{-s}^{s} \exp(-\mathbf{i}tr)\psi(z)dt$$

The procedure of quantile estimation is summarized as follows:

- Implement the discrete fast Fourier transformation (DFT) to approximate the density of r at every time point t:

  1. Let $N = 2^m$ with $m \in \mathbb{N}$ and define an equidistance grid over the integral interval $[-s, s]$ by setting $h = \frac{2s}{N}$ and the grid points $z_j = -s + j * h$ with $j = 0, \cdots, N$.
  2. Calculate the input of the DFT: $y_j = (-1)^j \psi(z_j^*)$ with $z_j^* = 0.5(z_j + z_{j+1})$ are the middle points. Notice that the characteristic function is time dependent.
  3. The density $f(r) = \frac{1}{2\pi} C_k \text{DFT}(y)_k$ with $C_k = \frac{2s}{N}(-1)^k \exp(-\frac{\mathbf{i}k\pi}{N})\mathbf{i}$ with $k = 0, \cdots, N-1$. We refer to [BDH05] and [MR04] for more details. The corresponding values of $r = -\frac{N\pi}{2a} + \frac{\pi k}{a}$.

- The cumulative density function and the quantile are then approximated based on the resulting density.

6.3 Covariance estimation with simulated data

In this section, the GHICA versus the DCC, are implemented to estimate covariance of simulated data. The dimension is set to be $d = 50$. The simulation study is designed to include structure shifts of covariance. To be more specific, the designed covariance changes among three matrices over time, one is an identity matrix denoted as Σ_1, meaning uncorrelatedness, and two symmetric and semi-positive defined matrices Σ_2 and Σ_3. (Here we first generate $d * d$ matrix U_1 whose elements are uniform random variables for Σ_2 and standard Gaussian variables for Σ_3, then calculate a new matrix $U_2 = U_1 * U_1'$ to guarantee the semi-positiveness. The elements $\Sigma(i,j)$ of the target matrix are calculated as $\Sigma(i,j) = U_2(i,j)/\sqrt{U_2(i,i)U_2(j,j)}$.) The ordered eigenvalues of these two matrices are displayed in Figure 6.2, by which the eigenvalues are distributed in $[5.92e-004, 3.779]$ (Σ_2) and $[0.002, 3.573]$ (Σ_3) respectively. The off-diagonal values span over $[-0.433, 0.468]$ in the first self-correlated matrix (Σ_2) and $[-0.447, 0.464]$ in the second one (Σ_3). Temporal stationarity is assumed to have a maximal length of 400 and a minimal length of 100. The structure shifts of the generated covariance are illustrated in Figure 6.3. The level of the shifts is either small with a shift from one self-correlated matrix (Σ_2 or Σ_3) to the identity matrix or contrariwise, e.g. at the point 700, or large with a change between the two self-correlated matrices, e.g. at the point 1800.

Furthermore, two distributional parameters μ and β of the standardized NIG innovations $\varepsilon_x(t)$ are set to be 0, meaning that the innovations are

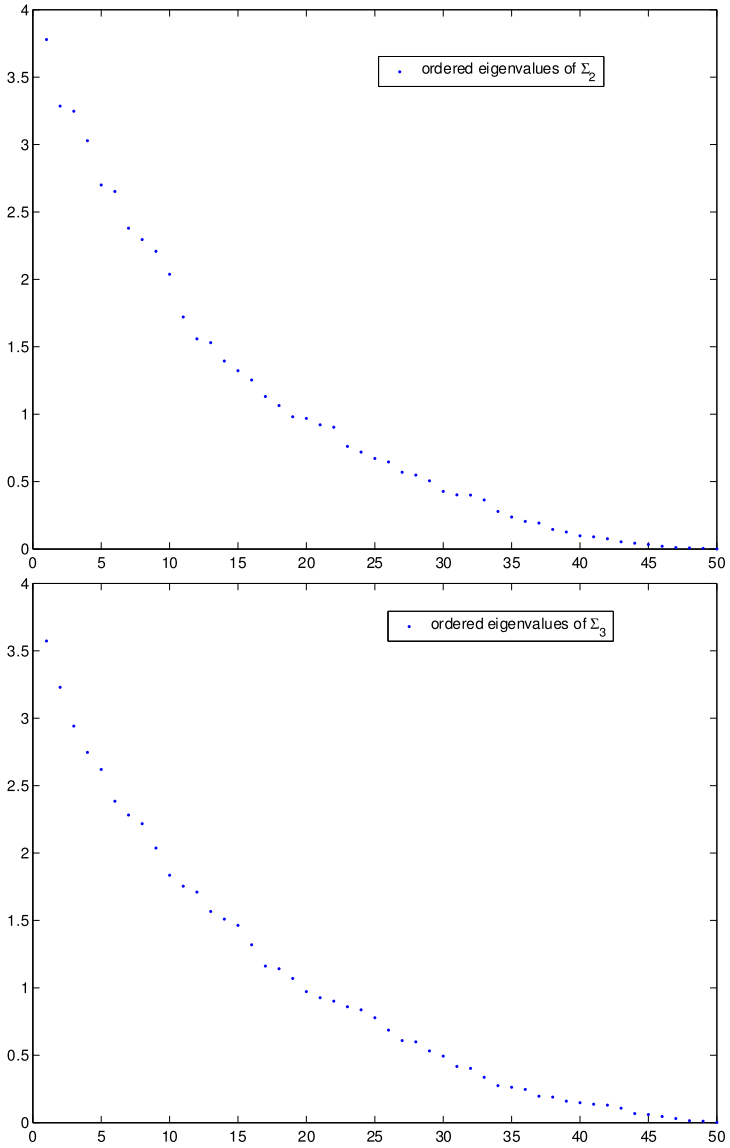

Figure 6.2: Ordered eigenvalues of the generated covariance matrices.

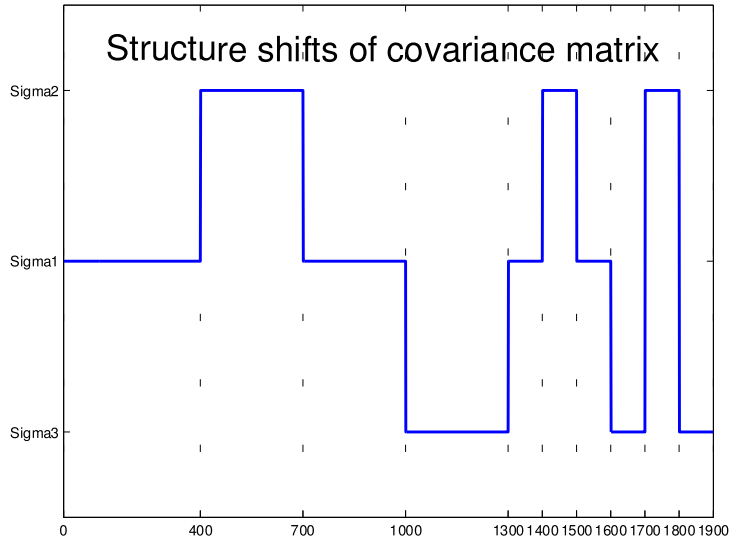

Figure 6.3: Structure shifts of the generated covariance through time. Notice that there are shifts among matrices not up-and-down movements.

centered around 0 and symmetric distributed, see [BNB81]. By doing so, the mean and variance of the NIG innovations only depend on α and δ:

$$\begin{aligned} \mathsf{E}(\varepsilon_x) &= \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}} = 0 \\ \mathrm{Var}(\varepsilon_x) &= \frac{\delta}{\sqrt{\alpha^2-\beta^2}} + \frac{\beta^2}{\delta^3\sqrt{\alpha^2-\beta^2}} = \frac{\delta}{\alpha} = 1 \end{aligned}$$

This result is used to generate the standardized innovations, by which $\alpha \sim U[1,2]$ is suggested by our experience on real data analysis and $\delta = \alpha$.

In the Monte Carlo simulation, we generate $d = 50$ NIG variables with the designed covariance and distributional parameters:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t).$$

The sample size is $T = 1900$ and the scenarios are repeated $N = 100$ times. The covariance matrix is estimated using the GHICA procedure and the DCC method respectively.

The GHICA method first converts the underlying series to ICs by a linear transformation:

$$x(t) = W^{-1}y(t) = W^{-1}D_y^{1/2}(t)\varepsilon_y(t),$$

by which the elements of $D_y(t)$ on the diagonal are estimated using the local exponential smoothing method. In the local exponential smoothing

estimation, we set the involved parameters $c = 0.01$, $a = 1.25$ and $p = 0.25$. The sequence of the smoothing parameters $\{\eta_k\}$ are $0.600, \cdots, 0.982$ with $K = 15$, based on the condition $(1-\eta_k)/(1-\eta_{k+1}) = a$ in (6.7). The first 300 observations are reserved as training set for the very beginning estimations, since the largest smoothing parameter used in this study corresponds to a window with 259 observations.

The covariance of $x(t)$ is calculated by the basic statistical property:

$$\Sigma_x(t) = W^{-1} D_y(t) W^{-1\top}$$

The DCC method assumes that the underlying series are Gaussian distributed. It decomposes the covariance matrix to a product of diagonal variance matrix and correlation matrix.

$$\Sigma_x(t) = D_x(t) R_x(t) D_x(t)^\top.$$

where $D_x(t)$ consists of the variances of $x(t)$ on the diagonal that are estimated in the GARCH(1,1) setup.

Figure 6.4 displays one realization of $\Sigma(2,5)$, i.e. the covariance of the second and fifth risk factors $x_2(t)$ and $x_5(t)$, based on one simulation data. The true values are 0.365 in Σ_2 and -0.124 in Σ_3. As expected, the GHICA estimates are sensitive to structure shifts through time. The DCC estimates, on the contrary, are over-smooth and slowly follow the shifts. Given more often shifts around the last hundreds of time points, the DCC estimates deliver less information on the movements. Recall that 100 points correspond to 4 months observations of daily returns. It is rational to surmise that structure shifts happen so often in the active financial markets, see [Mer73]. The similar estimation results are observed in the other elements of the covariance, which are eliminated here.

To measure the accuracy of estimation, ratio of absolute estimation error (RAE) of the estimates w.r.t. the true covariance are calculated pointwise.

$$\mathrm{RAE}(i,j) \quad = \quad \frac{\sum_{t=301}^{T} |\hat{\Sigma}^{\mathrm{GHICA}}_{(i,j)}(t) - \Sigma_{(i,j)}(t)|}{\sum_{t=301}^{T} |\hat{\Sigma}^{\mathrm{DCC}}_{(i,j)}(t) - \Sigma_{(i,j)}(t)|}$$

If $\mathrm{RAE}(i,j) \leq 1$, it means that the GHICA method reaches higher accuracy in the estimation of $\Sigma(i,j)$ than the DCC. To compare the general performance of these two methods in covariance estimation, we check the proportion of the RAEs among the 2500 $(d * d)$ elements that are smaller or equal to one, i.e. $\frac{\sum_i \sum_j \mathbf{1}(\mathrm{RAE}_{(i,j)\leq 1})}{d \times d}$ for $i, j = 1, \cdots, d$. Notice that the proportion with value of 0.5 indicates that half elements are better estimated by

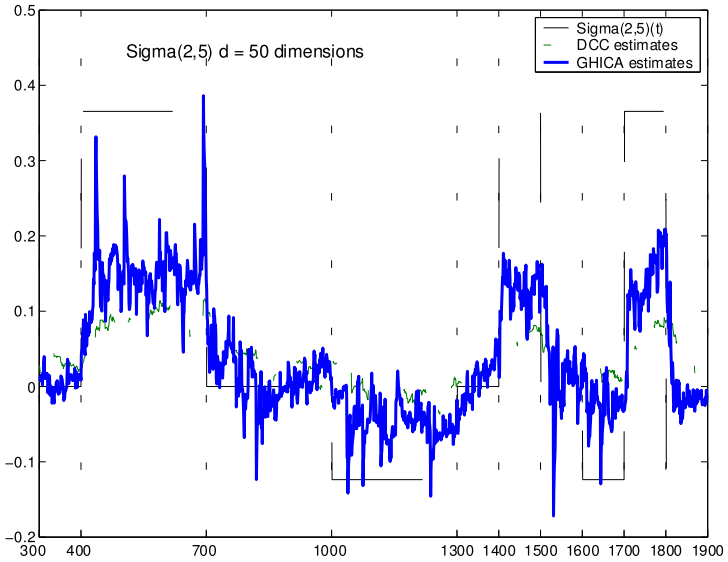

Figure 6.4: Realized estimates of $\Sigma(2,5)$ based on the GHICA and DCC methods. The generated data consists of 50 NIG distributed components.

using the GHICA and the other half are better done by the DCC. In other words, the considered methods have a comparable accuracy of estimation. Figure 6.5 displays the boxplot of the 100 proportions. The mean of the proportion is 0.4904 among the 100 simulations. It states that the DCC method performs a little bit better than the GHICA in the sense of accuracy. On the meanwhile, the GHICA method is fast and sensitive to structure shifts.

6.4 Risk management with real data

In this section, we implement the proposed GHICA method to calculate risk measures using real data sets: 20-dimensional German DAX portfolio and 7-dimensional exchange rate portfolio. The results are compared with those based on alternative risk management models. The data sets have been kindly provided by the financial and economic data center (FEDC) of the Collaborative Research Center 649 on Economic Risk of the Humboldt-Universität zu Berlin (http://sfb649.wiwi.hu-berlin.de). Before giving detailed description of the data sets, we analyze the risk measures from the viewpoints of regulatory, investors and internal supervisory. For more details, see Chapter 2.

Regulatory requirement: According to the Basel accord (1998), banks

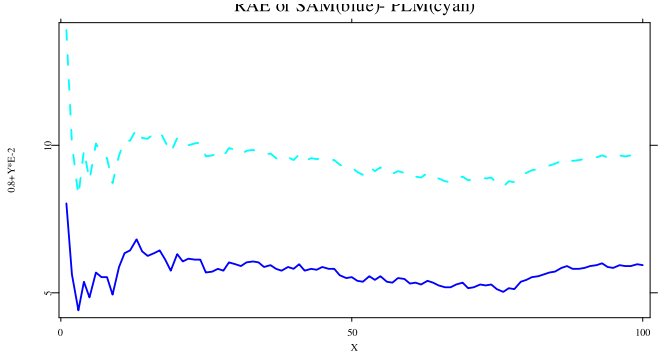

Figure 6.5: Boxplot of the proportion $\frac{\sum_i \sum_j \mathbf{1}(\text{RAE}_{(i,j)\leq 1})}{d\times d}$ for $i, j = 1, \cdots, d$. Here $d = 50$ and the proportions on the base of 100 simulations are considered.

are subject to a credit risk charge and allowed to use an internal model to measure their market risks, e.g. VaR as industry standard measure:

$$\text{VaR}_{t,1\%} = -\text{quantile}_{pr}\{r(t)\}.$$

where pr is the $h = 1$-day or $h = 5$-day ahead forecasted probability of the portfolio returns. The internal model is accepted if the exceptions over VaR at 1% probability spanning the last 250 days do not exceed 4. This regulatory rule actually suggests banks to control VaR at 1.6% probability and reserves risk charge based on the value. For simplification, we consider the empirical mean of VaR over the time horizon as the risk charge:

$$\text{Risk charge} = \text{mean}(\text{VaR}_t).$$

An internal model is particularly preferred by fulfilling the minimum regulatory requirement, i.e. with an empirical probability that is smaller or equal to 1.6%, and simultaneously requiring risk charge as small as possible.
Internal supervisory: It is interesting to choose a model that gives the empirical probability as close to the expected values as possible:

$$\hat{pr} = \frac{\text{No. exceedances}}{\text{No. total observations}}$$

Two extreme probabilities are considered, $pr = 1\%$ for regulatory reason and $pr = 0.5\%$ used by banks with AAA rating.
Investor: VaR is not a coherent risk measure. From the viewpoint of investors, for example, it is more interesting to measure the size of losses instead

of the frequency of losses. In accordance to this goal, ES has been considered as a better risk measure than the VaR:

$$\mathrm{ES} = \mathsf{E}\{-r_t | - r_t > \mathrm{VaR}_t\}.$$

Since investors suffer from a bankruptcy if the amount of losses is large, i.e. a large value of ES. They do prefer a small value of ES.

6.4.1 Data analysis 1: DAX portfolio

The primary target of the real data analysis is to compare the forecasting ability of the GHICA method with two alternatives, the RiskMetrics method under Gaussian distributional assumption and a modification with $t(6)$ distributional assumption (abbreviated as $t(6)$ method) in the market. The comparison is demonstrated based on 20 DAX stocks over a long time period, starting on 1974/01/02 and ending on 1996/12/30 (5748 observations). The return series are all centered around 0 and have heavy tails (kurtosis$>$ 3), the smallest correlation coefficient is 0.3654. Hypothetical German DAX portfolios are constructed with two static trading strategies $b(t) = b^{(1)} = (1/d, \cdots, 1/d)^\top$ and $b(t) = b^{(2)} \sim U[0,1]^d$. Such a simple portfolio construction eliminates the influence of strategy adjustments on the calculation. The portfolio returns are analyzed using the RiskMetrics or the $t(6)$ method. Here the unknown volatility process of the portfolio is estimated using the exponential smoothing method with $\eta = 0.94$:

$$\begin{aligned} r(t) &= b^\top x(t) = \sigma_r(t)\varepsilon_r(t) \\ \sigma_r^2(t) &= \{\sum_{m=0}^{M} \eta^m r^2(t-m-1)\}/(\sum_{m=0}^{M} \eta^m) \end{aligned}$$

where the truncated value M fulfills the condition $\eta^{(M+1)} \leq 0.01$. Notice that given a dynamic trading strategy, this simplification needs to repeatedly estimate the density of the time varying hypothetical portfolio returns, and further it often suffers from a low accuracy of estimation.

Figure 6.6 depicts the one day log-returns of the DAX portfolio with the static trading strategy $b(t) = b^{(1)}$. The VaRs from 1975/03/17 to 1996/12/30 at pr = 0.5% are displayed w.r.t. three methods, the GHICA, the RiskMetrics and the $t(6)$. The most volatile time period over $t \in [3300, 4300]$ is detailed in the bottom diagram. Recall that on the Monday, 19 October 1987, the worldwide downward jump of stocks happened. Dow Jones Industrial Average for example dropped by over 500 points. At this market quiver around $t = 3446$, the GHICA method exactly achieves the locations

		GHICA			RiskMetrics N(μ, σ^2)			Exponential smoothing $t(6)$		
$b(t)$	pr	p̂r	RC	ES	p̂r	RC	ES	p̂r	RC	ES
$b^{(1)}$	1%	0.55%	0.0264	0.0456	**1.18%**^{s}	**0.0229**^{r}	0.0279	0.40%	0.0292	**0.0269**^{i}
$b^{(1)}$	0.5%	**0.44%**^{s}	0.0297	**0.0472**^{i}	0.75%	0.0254	0.0317	0.23%	0.0345	0.0506
$b^{(2)}$	1%	0.59%	0.0265	0.0448	**1.03%**^{s}	**0.0231**^{r}	0.0288	0.38%	0.0294	**0.0406**^{i}
$b^{(2)}$	0.5%	**0.42%**^{s}	0.0298	**0.0476**^{i}	0.71%	0.0256	0.0315	0.21%	0.0347	0.0514
$b^{(1)}$	1%	0.83%	0.0550	0.0841	**1.15%**^{s}	**0.0481**^{r}	0.0602	0.19%	0.0665	**0.0833**^{i}
$b^{(1)}$	0.5%	**0.51%**^{s}	0.0612	**0.0939**^{i}	0.64%	0.0536	0.0683	0.09%	0.0784	0.1067
$b^{(2)}$	1%	**0.83%**^{s}	0.0554	**0.0828**^{i}	1.18%	**0.0488**^{r}	0.0613	0.16%	0.0673	0.0852
$b^{(2)}$	0.5%	**0.50%**^{s}	0.0617	**0.0943**^{i}	0.63%	0.0543	0.0676	0.07%	0.0794	0.1218

Table 6.1: Risk analysis of the DAX portfolios with two static trading strategies. The concerned forecasting interval is $h = 1$ (top) or $h = 5$ (bottom) days. The best results to fulfill the regulatory requirement are marked by ^{r}. The method preferred by investor is marked by ^{i}. For the internal supervisory, the method marked by ^{s} is recommended.

of extreme losses whereas the RiskMetrics and $t(6)$ methods over-react to them. Such over reactions induce large risk charges unnecessarily. On the other hand, it is observed that these two alternative methods give close forecasts to some extreme losses, e.g. around time points 4000 and 4500. As a result, the associating values of ES are small and satisfy the requirement of risk-averse investors.

Table 6.1 reports the risk measures based on the three methods. In general, the RiskMetrics is successful in fulfilling the minimal requirement of regulatory. The $t(6)$ method is preferred by investors who consider risk happened with 1% probability. The GHICA method performs better than the other two for internal supervisory and requirement of risk-averse investors who care the extreme risk happened with 0.5% probability.

6.4.2 Data analysis 2: Foreign exchange rate portfolio

In financial markets, traders adjust trading strategy according to information obtained. The GHICA is easily applicable to dynamic portfolios. We consider here 7 actively traded exchange rates, Euro (EUR), the US dollar (USD), the British pounds (GBP), the Japanese yen (JPY) and the Singapore dollar (SGD) from 1997/01/02 to 2006/01/05 (2332 observations). The foreign exchange rate (FX) market is the most active and liquid financial market in the world. It is realistic to analyze a dynamic portfolio with daily time varying trading strategy $b^{(3)}(t)$. The strategy at time point t relies on the

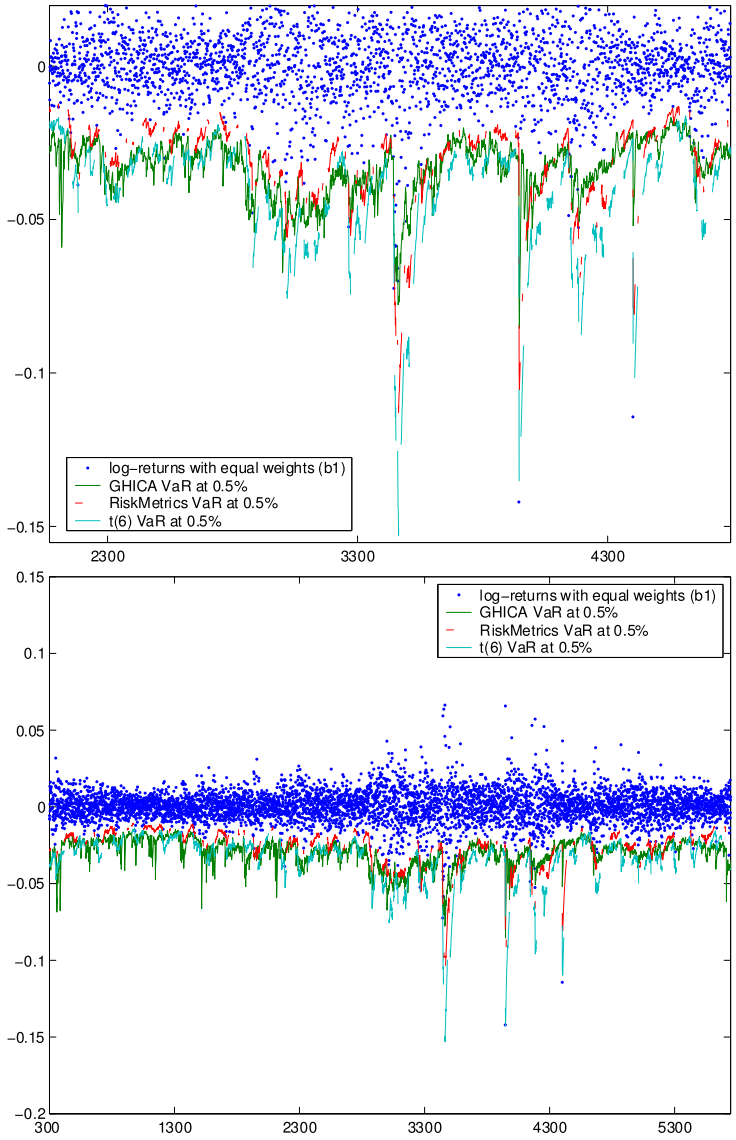

Figure 6.6: One day log-returns of the DAX portfolio with the static trading strategy $b(t) = b^{(1)}$. The VaRs are from 1975/03/17 to 1996/12/30 at pr = 0.5% w.r.t. three methods, the GHICA, the RiskMetrics and the $t(6)$. Part of the VaR time plot is enlarged and displayed on the bottom.

realized returns at $t-1$, the proportions of which w.r.t the sum of returns:

$$b^{(3)}(t) = \frac{x(t-1)}{\sum_{j=1}^{d} x_j(t-1)}$$

where $x(t) = \{x_1(t), \cdots, x_d(t)\}^\top$. Among these data sets, the returns of the EUR/SGD and USD/JPY rates are least correlated with the correlation coefficient 0.0071 whereas the returns of the EUR/USD and EUR/SGD rates are most correlated with the coefficient 0.6745. The resulting portfolio returns span over $[-0.7962, 0.7074]$.

The GHICA method is compared with an alternative method, abbreviated as DCCN, that applies the DCC covariance estimation under the Gaussian distributional assumption.

$$r(t) \;=\; b(t)^\top x(t) = b(t)^\top \Sigma_x^{(1/2)}(t)\varepsilon_x(t)$$

where $\varepsilon_x \sim \mathrm{N}(\mu, \Sigma_\varepsilon)$ with the diagonal covariance matrix Σ_ε. Notice that the quantile vector with pr-quantiles of individual innovation does not necessarily correspond to the pr-quantile of the portfolio return. Under the Gaussian distributional assumption, the standardized DCCN returns are theoretically cross independent and the Gaussian quantiles of the portfolio can be easily calculated. The dynamic mean, variance of the portfolio's returns have values of:

$$\begin{aligned} \mathsf{E}\{r(t)\} &= b(t)^\top \Sigma_x^{(1/2)}(t)\, \mathsf{E}\{\varepsilon_x(t)\} \\ Var\{r(t)\} &= b(t)^\top \Sigma_x^{(1/2)}(t)\; Var\{\varepsilon_x(t)\} \Sigma_x^{(1/2)\top}(t) b(t) \end{aligned}$$

The GHICA method in general presents better results than the DCCN. Except the value of ES at 1% level, the GHICA fulfills the requirements of regulatory, internal supervisory and investors, see Table 6.2. For $h = 1$ day forecasts, the DCCN gives although a closer VaR value to 1.6%, i.e. the ideal probability for regulatory, its risk charge with a value of 0.0494 is larger than that based on the GHICA, 0.0453. Therefore the GHICA is more favored in fulfilling the minimal regulatory requirement.

The two real data studies show that the GHICA method fulfills the minimal regulatory requirement by controlling the risk inside 1.6% level and requiring small risk charge, in particular satisfies the internal supervisory requirement by precisely measuring risk level as expected and favors the investors' requirement by delivering small size of loss. In summary, the GHICA method is not only a fast procedure given either static or dynamic portfolios but also produces better results than several alternative risk management methods.

			GHICA			DCCN		
h	$b(t)$	pr	p̂r	RC	ES	p̂r	RC	ES
1	$b^{(3)}(t)$	1%	**1**.28%^{s}	**0**.0453^{r}	0.0778	1.59%	0.0494	**0**.0254^{i}
	$b^{(3)}(t)$	0.5%	**0**.59%^{s}	0.0493	**0**.1944^{i}	0.94%	0.0547	0.0289
5	$b^{(3)}(t)$	1%	**1**.53%^{s}	**0**.0806^{r}	**0**.2630^{i}	4.17%	0.0993	0.1735
	$b^{(3)}(t)$	0.5%	**0**.79%^{s}	0.1092	**0**.2801^{i}	3.44%	0.1100	0.1389

Table 6.2: Risk analysis of the dynamic exchange rate portfolio. The best results to fulfill the regulatory requirement are marked by ^{r}. The recommended method to the investor is marked by ^{i}. For the internal supervisory, we recommend the method marked by ^{s}.

Bibliography

[ABCD05] Andersen, T.G.; Bollerslev, T.; Christoffersen, P.F.; Diebold, F.X.: Volatility and correlation forecasting. In: Elliott, G.; Granger, C.W.J.; Timmermann, A., editors, *Handbook of Economic Forecasting*. Amsterdam: North-Holland, 2005.

[ABDL01] Anderson, T.G.; Bollerslev, T.; Diebold, F.X.; Labys, P.: The distribution of realized exchange rate volatility. In: *Journal of the American Statistical Association*, pp. 42–55, 2001.

[BC64] Box, G.E.P.; Cox, D.R.: An analysis of transformations. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252, 1964.

[BDH05] Borak, S.; Detlefsen, K.; Härdle, W.: FFT-based option pricing. In: Cizek, P.; Härdle, W.; Weron, R., editors, *Statistical Tools for Finance and Insurance*. Springer Verlag, 2005.

[BN77] Barndorff-Nielsen, O.: Exponentially decreasing distributions for the logarithm of particle size. In: *Proceedings of the Royal Society of London*, volume A 353:pp. 401–419, 1977.

[BN97] Barndorff-Nielsen, O.: Normal inverse gaussian distributions and stochastic volatility modelling. In: *Scandinavian Journal of Statistics*, volume 24:pp. 1–13, 1997.

[BNB81] Barndorff-Nielsen, O.E.; Blæsild, P.: Hyperbolic distribution and ramifications: Contributions to theory and applications. In: Taillie, C.; Patil, P.G.; Baldessari, A.B., editors, *Statistical Distributions in Scientific Work*, volume 4, pp. 19–44. D. Reidel, 1981.

[BNS01] Barndorff-Nielsen, O.E.; Shephard, N.: Modelling by lévy processes for financial econometrics. In: Barndorff-Nielsen, O.;

Mikosch, T.; Resnik, S., editors, *Lévy Processes : Theory and Applications*. Birkhauser Boston, 2001.

[Bol86] Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. In: , volume 31:pp. 307–327, 1986.

[Bol90] Bollerslev, T. P.: Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. In: *Review of Economics and Statistics 72*, pp. 498–505, 1990.

[Bol95] Bollerslev, T.: Generalied autoregressive conditional heteroskedasticity. In: Engle, R., editor, *ARCH, selected readings*, pp. 42–60. Oxford University Press, 1995.

[Bol01] Bollerslev, T.: Financial econometrics: Past developments and future challenges. In: , volume 100:pp. 41–51, 2001.

[BS01] Bibby, B. M.; Sørensen, M.: *Hyperbolic Processes in Finance, Technical Report 88*. University of Aarhus, Aarhus School of Business, 2001.

[BS06] Belomestny, D.; Spokoiny, V.: Spatial aggregation of local likelihood estimates with applications to classification. In: *WIAS Preprint*, 2006.

[BW92] Bollerslev, T.; Woolridge, J.M.: Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. In: *Econometric Reviews*, volume 11:pp. 143–172, 1992.

[BW98] Back, A.D.; Weigend, A.S.: A first application of independent component analysis to extracting structure from stock returns. In: *International Journal of Neural Systems*, volume 8:pp. 473–484, 1998.

[CFS03] Cheng, M.Y.; Fan, J.; Spokoiny, V.: Dynamic nonparametric filtering with application to volatility estimation. In: Akritas, M.G.; Politis, D.N., editors, *Recent advances and trends in nonparametric statistics*, pp. 315–334. Elsevier, 2003.

[CHJ05] Chen, Y.; Härdle, W.; Jeong, S.O.: Nonparametric risk management with generalized hyperbolic distributions. In: *SFB 649, Discussion paper 2005-001*, 2005.

[Chr98] Christoffersen, P. F.: Evaluating interval forecast. In: *International Economic Review*, volume 39:pp. 841–862, 1998.

[CHS6a] Chen, Y.; Härdle, W.; Spokoiny, V.: Portfolio value at risk based on independent components analysis. In: *Journal of Computational and Applied Mathematics, forthcoming*, 2006a.

[CHS6b] Chen, Y.; Härdle, W.; Spokoiny, V.: Ghica - risk analysis with gh distributions and independent components. In: *SFB 649, Discussion paper*, 2006b.

[CS06] Chen, Y.; Spokoiny, V.: Local exponential smoothing with applications to volatility estimation and risk management. In: *working paper*, 2006.

[CT91] Cover, T.M.; Thomas, J.A.: *Elements of information theory*. Wiley, 1991.

[DF79] Dickey, D.A.; Fuller, W.A.: Distribution of the estimators for autoregressive time series with a unit root. In: *Journal of the American Statistical Association*, volume 74:pp. 427–431, 1979.

[Dia91] Diamond, D.W.: Debt maturity structure and liquidity risk. In: , volume 106:pp. 709–737, 1991.

[DJK+02] Duann, J.R.; Jung, T.P.; Kuo, W.J.; Yeh, T.C.; Makeig, S.; Hsieh, J.C.; Sejnowski, T.J.: Single-trial variability in event-related bold signals. In: *NeuroImage*, volume 15:pp. 823–835, 2002.

[Duf99] Duffee, G.R.: Estimating the price of default risk. In: , volume 12:pp. 197–226, 1999.

[EGS86] Eaton, J.; Gersovitz, M.; Stiglitz, J.E.: The pure theory of country risk. In: , volume 30:pp. 481–514, 1986.

[EK95a] Eberlein, E.; Keller, U.: Hyperbolic distributions in finance. In: *Bernoulli*, volume 1:pp. 281–299, 1995.

[EK95b] Engle, R.; Kroner, F.: Multivariate simultaneous generalized arch. In: *Econometric Theory 11*, pp. 122–150, 1995.

[EKK03] Eberlein, E.; Kallsen, J.; Kristen, J.: Risk management based on stochastic volatility. In: *Journal of Risk*, volume 5:pp. 19–44, 2003.

[EMS99] Embrechts, P.; McNeil, A.; Straumann, D.: Correlation: pitfalls and alternatives. In: *Risk*, volume 12:pp. 69–71, 1999.

[Eng82] Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. In: , volume 50:pp. 987–1007, 1982.

[Eng95] Engle, R. F.: *Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, ARCH.* Oxford University Press, 1995.

[Eng02] Engle, R.: Dynamic conditional correlation - a simple class of multivariate garch models. In: *Journal of Business and Economic Statistics, 20(3)*, pp. 339–350, 2002.

[EP02] Eberlein, E.; Prause, K.: The generalized hyperbolic model: financial derivatives and risk measures. In: Geman, H.; Madan, D.; Pliska, S.; Vorst, T., editors, *Mathematical Finance-Bachelier Congress 2000*. Springer Verlag, 2002.

[ES01] Engle, R.F.; Sheppard, K.K.: Theoretical and empirical properties of dynamic conditional correlation multivariate garch. In: *NBER Working Paper 8554*, 2001.

[Fam65] Fama, E.: The behavior of stock prices. In: , volume 38:pp. 34–105, 1965.

[FHH04] Franke, J.; Härdle, W.; Hafner, C.: *Statistics of Financial Markets*. Springer-Verlag Berlin Heidelberg New York, 2004.

[FHW95] Fan, J.; Heckman, N.E.; Wand, M.P.: Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. In: *Journal of the American Statistical Association*, pp. 477–489, 1995.

[Flu98] Flury, B.: *Common Principal Components and Related Multivariate Models*. John Wiley & Sons, Inc., 1998.

[HHS03] Härdle, W.; Herwartz, H.; Spokoiny, V.: Time inhomogeneous multiple volatility modelling. In: *Journal of Financial Econometrics*, volume 1:pp. 55–95, 2003.

[HKO01] Hyvärinen, A.; Karhunen, J.; Oja, E.: *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.

[HMSW04] Härdle, W.; Müller, M.; Sperlich, S.; Werwatz, A.: *Nonparametric and Semiparametric Models*. Springer Verlag, 2004.

[HO99] Hyvärinen, A.; Oja, E.: Independent component analysis: Algorithms and applications. In: *Neural Networks*, volume 13:pp. 411–430, 1999.

[HRS95] Harvey, A.; Ruiz, E.; Shephard, N.: Multivariate stochastic variance models. In: Engle, R., editor, *ARCH, selected readings*, pp. 256–276. Oxford University Press, 1995.

[HS03] Härdle, W.; Simar, L.: *Applied Multivariate Statistical Analysis*. Springer-Verlag Berlin Heidelberg New York, 2003.

[Hul97] Hull, J.C.: *Options, futures, and other derivatives*. Prentice-Hall International, Inc., 1997.

[Hyv98] Hyvärinen, A.: *New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit*, pp. 273–279. MIT Press, 1998.

[IM98] Iyengary, S.; Mazumdar, M.: A saddle point approximation for certain multivariate tail probabilities. In: *SIAM Journal on Scientific Computing*, volume 19:pp. 1234–1244, 1998.

[Jas01] Jaschke, S.: Quantile-var is the wrong measure to quantify market risk for regulatory purposes. In: , 2001.

[JJ02] Jaschke, S.; Jiang, Y.: Approximating value at risk in conditional gaussian models. In: Härdle, W.; Kleinow, T.; Stahl, G., editors, *Applied Quantitative Finance*. Springer Verlag, 2002.

[Jor01] Jorion, P.: *Value at Risk*. McGraw-Hill, 2001.

[JS87] Jones, M.C.; Sibson, R.: What is projection pursuit? In: *Journal of the Royal Statistical Society, A 150(1)*, pp. 1–36, 1987.

[KMR03] Kahn, C.M.; Mcandrews, J.; Roberds, W.: Settlement risk under gross and net settlement. In: , volume 35:pp. 591–608, 2003.

[Man63] Mandelbrot, B.: The variation of certain speculative prices. In: *The Journal of Business*, volume 36:pp. 394–419, 1963.

[Mer73] Merton, R.C.: Theory of rational option pricing. In: *The Bell Journal of Economics and Management Science*, volume 4:pp. 141–183, 1973.

[MN89] McCullagh, P.; Nelder, J.A.: *Generalized linear models*. Chapman and Hall London, 1989.

[MR04] Menn, C.; Rachev, S.T.: Calibrated FFT-based density approximations for α-stable distributions. In: , 2004.

[MS00] Mikosch, T.; Stărică, C.: Change of structure in financial time series, long range dependence and the garch model. In: , 2000.

[MS04a] Mercurio, D.; Spokoiny, V.: Statistical inference for time inhomogeneous volatility models. In: *The Annals of Statistics*, volume 32:pp. 577–602, 2004.

[MS04b] Mercurio, D.; Spokoiny, V.: Volatility estimation via local change point analysis with applications to value-at-risk. In: *submitted*, 2004.

[Pag96] Pagan, A.: The econometrics of financial markets. In: , volume 3:pp. 15–102, 1996.

[PG03] Poon, S.H.; Granger, C.W.J.: Forecasting volatility in financial markets: A review. In: *Journal of Economic Literature*, volume XLI:pp. 478–539, 2003.

[Pra99] Prause, K.: The generalized hyperbolic model: Estimation, financial derivatives and risk measures. In: *dissertation*, 1999.

[PS06] Polzehl, J.; Spokoiny, V.: Propagation-separation approach for local likelihood estimation. In: *Probability Theory and Related Fields*, pp. 335–362, 2006.

[PTVF92] Press, W.T.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P.: *Numerical Recipes in C*. Cambridge University Press, 1992.

[RRK02] Ristaniemi, T.; Raju, K.; Karhunen, J.: Jammer mitigation in ds-cdma using independent component analysis. In: *In Proc. of the IEEE Int. Conf. on Communications*, 2002.

[Spo98] Spokoiny, V.: Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. In: *The Annals of Statistics*, volume 26:pp. 1356–1378, 1998.

[Spo06] Spokoiny, V.: *Local parametric methods in nonparametric estimation*. Springer-Verlag Berlin Heidelberg New York, 2006.

[VW97] Viola, P.; Wells, WM III: Alignment by maximization of mutual information. In: *International Journal of Computer Vision*, pp. 137–154, 1997.

[Wil00] Wilmott, P.: *Paul Wilmott on quantitative finance*. John Wiley and Sons Ltd., 2000.

Acknowledgement

I would like to express my deep gratitude and respect to my advisers, prof. Dr. Wolfgang Härdle and prof. Dr. Vladimir Spokoiny, who opened me the door of research in quantitative finance, adaptive data analysis and many others, from whom I have learned not only the modern methods but also their working style and altitude towards science, namely responsibility and precious. I would never forget that they have motivated, enlightened and advised me during my research. I also want to thank my friends, coauthors and colleagues at the Institute of Statistics of Humboldt University in Berlin and at the Wierstrass Institute for Applied Analysis and Stochastics, Denis Belomestny, Michal Benko, Szymon Borak, Zdeněk Hlávka, Seok-Oh Jeong, Karsten Tabelow and many others, thank for their cooperation and friendship these years. Finally, I give my special thanks to my dear husband, Yanwu Wang, whose support was extremely valuable and significantly contributed to the realization of this thesis.

Ying Chen
September 2006, Berlin

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Ying Chen
26 September 2006